A New Teleconference System by GigaEther Omni-directional Video Transmission for Healthcare Applications

Hirokazu Ookuzu, Yosuke Sato, Koji Hashimoto and Yoshitaka Shibata
Faculty of Software and Information Science, Iwate Prefectural University, Japan
g031d021@edu.soft.iwate-pu.ac.jp, g031b074@edu.soft.iwate-pu.ac.jp, hashi@iwate-pu.ac.jp, shibata@iwate-pu.ac.jp

Abstract

In this paper, a new teleconference system by using Gigabit Ethernet based omni-directional camera system and flexible middleware system in realtime and bi-directional video transmission is introduced. The omni-directional image has more advantages than conventional one-directional camera in that it provides a wider view than a single directional camera and able to realize flexible teleconferencing even between remotely separated small rooms. System architecture and functions of the middleware for high-definition omni-directional video control and effective video transmission system using Windows Media Video(WMV) format [4] are precisely described. QoS control function is also introduced to dynamically control the frame rate of video depending on the changes of network and CPU load. The prototype system of a teleconference is constructed to apply for remote healthcare education and evaluate the performance of our suggested high-definition omni-directional video system. Through the performance evaluation of the prototyped system, we could verify the usefulness of our proposed system.

1. Introduction

Recently high-power personal computers have been commonly used for teleconference system with high quality audio and video. So far, since those teleconference systems usually use one directional video camera with limited angle, it is very difficult to realize the teleconference with many participants particularly in a small room and to capture wide range of the inside area in real time although this is particularly important for healthcare system where medical doctors or mental counselors diagnose precisely have to remotely consult and diagnose many patients at the same time.

In this paper, in order to overcome those problems, we introduce a new teleconferencing system and transmission middleware system by omni-directional camera which can capture a view of 360 degree angle in realtime. In the middleware system, MidField system [1] which we have developed so far is used for video stream transmission over IP network. Then the quality of the panorama and expanded video images can be controlled and guaranteed by considering the computer and network resources. A prototype system of bi-directional teleconference system using gigabit Ethernet based on omni-directional camera (GigaEther camera) is constructed to use for healthcare applications and its function and performance are evaluated. As result, we could verify the usefulness of our proposed system.

In the followings, a system configuration of bi-directional teleconference system based on the omni-directional image is precisely described in section 2. A middleware system and its system architecture of suggested TV conference are explained in section 3. The Midfield system [1] which realizes a basic function of video stream transmission over IP network is explained in section 4. The video stream control method and QoS control method for teleconference system are precisely explained in section 5. The implementation and prototype system for healthcare applications are introduced in section 6.

2. System Configuration

The system configuration of our suggested teleconference system as shown in Fig.1 is consisted of server and client PCs. The server using GigaEther video camera (2048x2048, 15fps) with a omni-directional lens, so called PAL lens, as shown in Fig. 2 captures 360 degree ring-shaped image and audio sound. Those images are captured on the server PC through gigabit Ethernet interface while audio sound is taken by omni-directional microphone with echo canceling function.

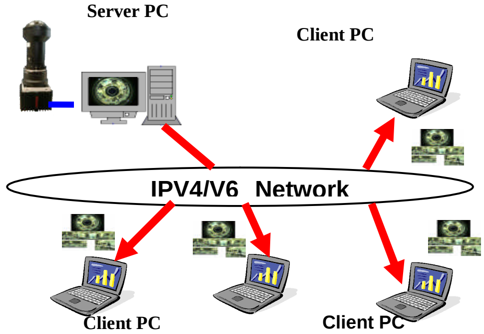

Fig. 1 System Configuration

Fig. 2 GigaEther Omni-Directional Video Camera

In typical, at the server, the captured ringed image and audio sound are transmitted over IPv4 or IPV6 based network as RTP video streams to the client. At the client PC, the ringed image is converted to the equivalent panorama image, expanded and displayed on the screen together with audio sound. Then any specific area selected by client user is calibrated and output on monitor..

When the network bandwidth is limited and not enough to transmit the ringed video stream or the client PC power is not enough to process the ringed image, the ringed image is compressed using various image coding methods and sent to the client to reduce the required bandwidth using M-JPEG, MPEG-2/4, WMV etc..

In order to realize bi-directional teleconference, a set of client and server are simultaneously used at each location. Since omni-directional video provides 360 degree image, the user can view wide range of captured image and freely select any specific part he wants to look st and can expand the desired part of the panorama image by mouse operations. The user's viewpoint can freely control in both vertical and horizontal directions and zooming as if he could control the angle of an ordinal one-directional camera.

The captured ringed image is encoded and decoded to the M-JPEG, MPEG-2/4 or WMV compressed video formats on either the server or client depending on the CPU loads and network resource condition and transmitted using MidField which is a middleware system to transmit various formatted video streams over IP network and will be precisely explained its functionality in the following section.

3. System Architecture

The system architecture of the proposed omni-directional TV conference system consists of five layers including Application layer, Omni-directional development layer including Panorama, Expanded and Calibrated Image sub-layers, Synchronization layer, Session layer and Midfield layer as shown in Fig. 3.

In the application layer, sampling the ringed video image and audio sound and displaying those on the monitor and speaker for TV conference are performed. In the omni-directional development layer, image development from the ringed image to the panorama, expanded and calibrated images are processed.

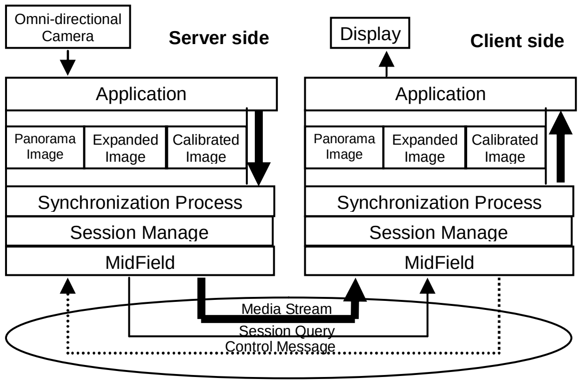

Fig. 3 System Architecture

In the synchronization layer, audio and the panorama, expanded and calibrated image are synchronized. In the session layer, a TV conference session by those images and audio is managed. In the Midfield layer, those videos and audio are transmitted as a stream in real time over IP network.

4. MidField System

As shown in Fig. 4, the MidField system is located between the application and the transport layer. The system is constructed by 3 layers and 4 vertical planes and offers multimedia communication functions to the application layer. Stream Plane is constructed by synchronization, data transform and media flow control layer, and performs multimedia stream processing. Session Plane performs management of communication sessions. System Plane monitors network traffic and CPU rate in the local host, and performs admission tests for QoS requirements from system user. Event Process Plane processes various events that are created in the system.

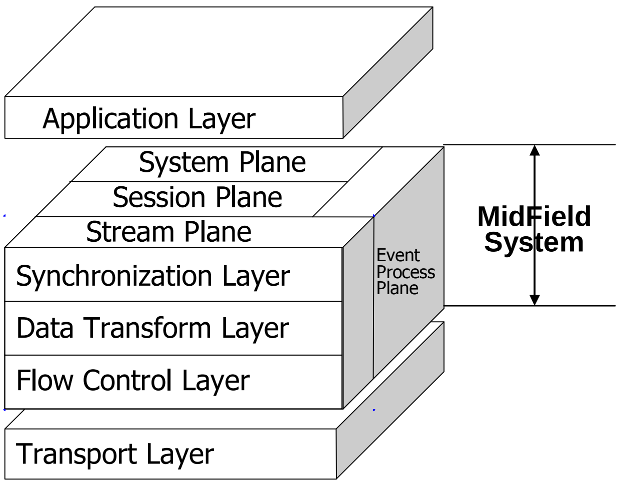

Fig. 4 MidField System

In the Session Plane of MidField System, information of both participants and media streams is handled. It should be considered that some new functions for handling this information are required by various session types. Therefore if both these information and the handling functions can transfer to intermediate node that performs to transcode, the system will have flexibility for updating new management functions for new session types. In addition, if there are no suitable functions to process media in local computer system, the system should have some functions to get required media processing functions from the other systems.

The system is able to construct intercommunication environment on computer networks dynamically according to the environment of users and QoS requirements from users. Fig. 5 shows an abstract of MidField Session. A MidField Session consists of at least one multicast session and offers peer-to-peer communication to system users. In Fig. 5, three system users (MF1, MF2 and MF3) join to a MidField Session. At this point, both MF1 and MF2 have enough communication environments to transmit DV stream. On the other hand, MF3 can't transmit DV stream because the environment doesn't have enough bandwidth and computing power to handle DV stream. Therefore MF3 requests to use MPEG4 stream to join the MidField Session. In such case, MidField System locates required transcoding functions into suitable node on computer networks to communicate with each other. In this example, both MF4 and MF5 transcode from DV stream into MPEG4. By using transcoding functions, MF3 can join to the MidField Session.

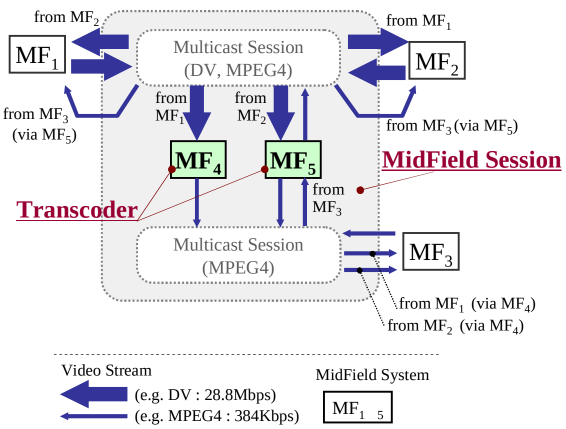

Fig. 5 MidField Session

5. Video Process and Control of Streaming

5.1 Omni-directional Video

There are several kinds of omni-directional cameras [2]. In our research, we apply a combination of GigaEther video camera and omni-directional lens, so called PAL lens as previously shown in Fig. 2. Using the omni-directional camera, the ring-shaped image with the 360 degree angle can be taken into the server PC through gigabit Ethernet. The captured ringed image includes 2048x2048 pixels as a resolution and 15 fps as a frame rate from -20 to 40 degree of a vertical view angle using GigaEther video camera. This ringed image is processed by the middleware on the server to the panorama image to be easily understood by user. Since the panorama image by PAL lens can not be zoomed in optical way, the any specific part of the panorama image is expanded by digital zooming method. However, digital zooming makes the image coarse as the enlargement rate increases. Therefore, the enlargement rate is limited within 3 to 5 times.

The image calibration for the expanded image from the selected area is carried out to remove the image distortion by PAL lens and to reconstruct the corrected image as if it were taken from the conventional one-directional camera.

5.2 Variation of omni-directional image

The image taken through the PAL lens is called ringed circle image. Three kinds of images shown below from this ringed circle image are generated, as shown in Fig.6.

- Panorama image which is converted form the ringed circle image taken through the PAL lens of 360 degree.
- Expanded image which is clipped and expanded at arbitrary position from the panorama image.
- Calibrated image which is removed distortion of omni-directional image from the expanded as taken from the regular one-directional camera.

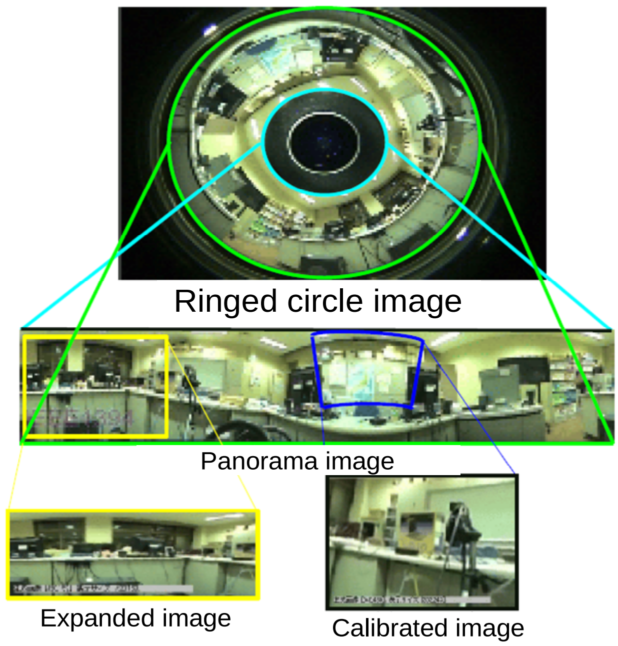

Fig. 6 Variation of omni-directional images

5.3 Omni-directional image development process

The development process from the ringed image to a panorama image is carried out in the following procedure as shown in Fig. 7.

1) Making a coordinate conversion table
2) Calibration of ringed image
3) Image development of panorama, expanded and calibrated images
4) Image interpolation

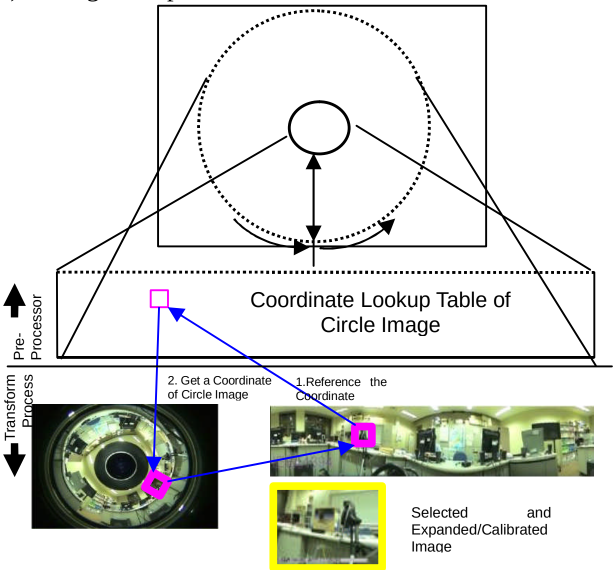

Fig. 7 Omni-directional Image Processing

At the 1st step, since the image development of the ringed image is realized based on polar coordinate conversion, the computation time of image processing, particularly for addressing of memory location for a panorama image is very large. For this reason, the coordinate conversion table form polar axis to x-y axis is produced based on the size of ringed image and visual angle of lens in advance to reduce the computation time for image conversion.

At the 2nd step, since the ringed image with 2048x2048 pixel resolution is captured by GigaEther video camera and sent to server PC through gigabit Ethernet in row data. Then ringed image is processed to convert to a panorama image with 3000x500 pixel resolution by the DirectShow filter in MS Window operating system,

At the 3rd step, by referring the coordinate conversion table, the panorama, expanded and calibrated images for the specified area can be extracted from the ringed circle image.

At the 4th step, the extracted image is interpolated to be smoothly displayed on the monitor. The surrounding 4 neighbor pixels of any points are interpolated by linear filtering or 2 dimensional spatial FFT or sub-band filtering methods [3]. By repeating from the 1st step through the 4th step, the original raw ringed image can be changed to desired images.

5.4 Stream Control

In our system, there are two streaming control methods between client and sever, including client based type and server based type. In the client based type, which is normally used, as shown in Fig.8, the ringed images captured on the server side are multicasted to the client side using Midfield. On the client side, the multicasted ringed image is received at client and processed to the panorama image, the expanded image and calibrated image for the selected area depending on the user's selection.

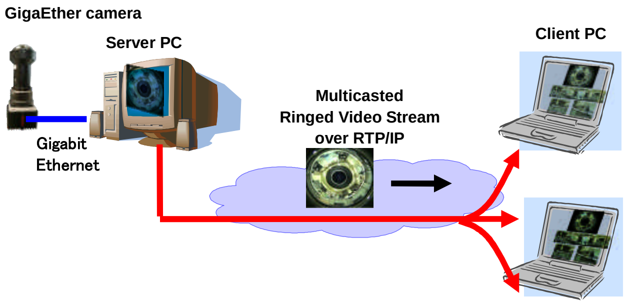

Fig. 8 Client Based Type

Using this client based type system, the ringed image is freely processed on the client and any area in the panorama image can be expanded on each client. On the other hand, since the ringed image with transcoded format have to be processed on the client side, more powerful CPU is required for client PC.

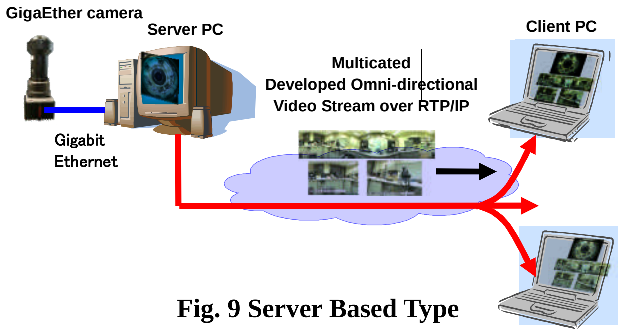

Fig. 9 Server Based Type

In the server based type, on the other hand, as shown in Fig. 9, the ringed image captured form GigaEther video camera on the server side is first converted to the equivalent panorama image, expanded and calibrated for the specific area selected by some client and transcoded to various video formats including M-JPEG, MPEG-2/4, WMV by Midfield, and then multicasted to the client.

The client receives the multicaseted stream and decodes to display on the monitor. Using this server based type, an image conversion from the ringed image to panorama image can be skipped on the client. Therefore, the relatively low powered PC can be used to obtain on the client side. However, since all of image conversions and format generation are executed on the server, the CPU load on the server increases. The control of the viewpoint to all of the clients is also limited.

By combing those two methods, the more flexible and proper omni-directional video stream transmission according to the user's environment can be attained.

5.5 QoS Control

In addition to those stream control methods, QoS control based on the client and server CPU load and network traffic is introduced in this system. By introducing QoS control, more proper and precise quality of video and audio can be provided depending on the user's environment. When the CPU load on the client and/or server increases, the frame rate of the video stream degreases. When the number of the high quality video streams such as WMV streams increases, the packet loss may increases. Since the load condition such as CPU occupation rate or CPU load average on the client and server can be observed, the image size and frame rate of the omni-directional image can be controlled.

On the other hand, the required network bandwidth varies depending on the size of the image and its video formats and compression rate. Therefore, by controlling the size and frame rate of the transmitted video stream according to the dynamically available network bandwidth, the omni-directional video image can be adjusted to user's environment.

6. Prototyped System and Evaluation

In order to verify the effects of our proposed teleconference system, we constructed a prototype system and evaluated its functional and performance through two different healthcare applications, namely remote mental healthcare groupwork system and remote nursing practice system as shown in Fig. 10. The prototyped teleconference system consists of server PC with GigaEther omni-directional video camera with, four monitors and omni-directional microphone, and client PCs with HDV camera and monitor are and connected to the campus VLAN network through gigabit Ethernet switching hub. The specification of the both server and client PCs is shown in shown in Table 1.

Table 1 Protype Specification

Specification	Server	Client
CPU	Xeon quadcore 3.2GHx 2CPU	Pentium4 3.2GHz
Main Memory	8GB	1.5GB
Network	1Gbps	1Gbps

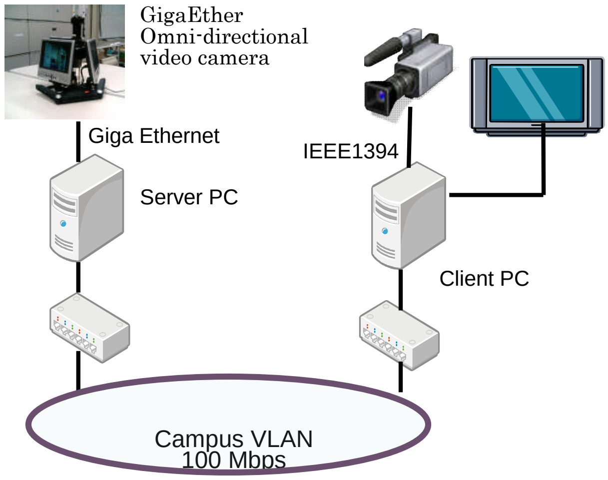

Fig. 10 Prototype System

The middleware system of omni-directional video transmission was developed using C++ (Microsoft Visual C++ .NET 2003) language and DirectX 9.0b (DirectShow) for video and audio processing and their managements modules. Since our prototype provides bi-directional teleconference facility, bi-directional audio and video stream transmissions were carried out to evaluate its performance.

As the video format of omni-directional image, the ringed image from GigaEther video camera was converted to panorama image at server PC, encoded to WMV format image, and then multicasted to the client PC. The received panorama image was decoded, expanded and calibrated at client PC, The result of performance evaluation is shown in Table 2.

Table 2 Result of Performance Evaluation

Iterms	Server	Client
Video Format	Omni-directional	DV, HDV
Pixel Resolution	2000x350	760x480, 1440x1080
Codec	WMV	WMV
Req. BW	4 Mbps	4 Mbps
Frame Rate	11 fps	29 fps
Ave. CPU load	30 %	30 %

From this performance evaluation, the frame rate of omni-directional panorama image at server side was 11 fps which was almost enough for users to smoothly keep teleconference for healthcare application while the frame rate of DV and HDV was almost perfectly attained.

5.1. Remote Mental Healthcare Groupwork

Next we actually applied our teleconference system for remote mental healthcare groupwork application. In the groupwork where the mental participants of a local government including the patient, his colleagues, family come together around circle table and discuss how the patient can resolve his problem and smoothly recover to work. Through the groupwork process, mental doctor and counselors remotely support the flow of the progress by checking all of the participant's faces, behaviors, opinions and counseling individually from remote site. In order to support this mental healthcare groupwork, omni-directional video was set at the center of the circler table as shown in Figure. 11. Thus, the omni-directional image of all of the participants can be captured at the same time and sent and developed into panorama images at the remote site where the mental doctor checks the panorama images on the display. Reversely his image and voice are sent to the groupwork site. Those images can be stored and replayed later on by simple operations. Through this experiment, groupwork could be effectively carried out.

Figure. 1 Remote Mental Healthcare Groupwork

5.2. Remote Nursing Practice System

In the course of nursing school, practical training is compulsory to qualify as nurse. In order to carry out the practical training, multiple omni-directional camera were installed in the ward where many nursing students are receiving practical training from the medical instructor located at the remote hospital by checking those students' medial operations and treatments for the imitated patients and advising individually using the teleconference system. Thus, using this system, the instructor could wisely view the students in the ward and precisely check whether the students can correctly treat and operate the patients even though there are many students in the wide ward.

7. Conclusions

In this paper, a new teleconference system by using Gigabit Ethernet based omni-directional camera system based on flexible middleware system for realtime and bi-directional video transmission was introduced. The omni-directional image has more advantages than conventional one-directional camera in that it provides a wider view than a single directional camera and able to realize flexible teleconferencing even between remotely separated small rooms.

A middleware for High Definition Omni-directional Video Transmission over IP Network was also introduced and its system architecture and functions are precisely described. The video stream transmission of the ringed image over gigabit network based on the Midfield which was previously developed for omni-directional video and audio was provided. The teleconference system as a prototyped system was constructed to evaluate the performance and applied for two healthcare applications. As a result, our system could provide satisfied performance which was almost equivalent to the one by the conventional one directional teleconference system with higher resolution.

Currently we are optimizing the omni-directional video processing to improve performance and implementing QoS control function according to the CPU load on both client and server and network traffic.

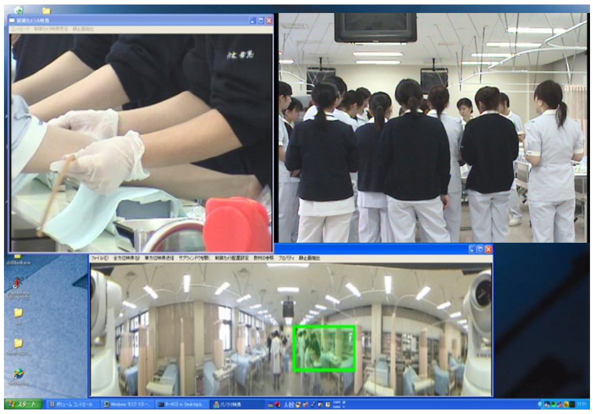

Figure. 2 Remote Nursing Practice

References

[1] Koji Hashimoto, Yoshitaka Shibata, "Design of a Middleware System for Flexible Intercommunication Environment, " IEEE Proc. on Advanced Information Networking and Applications, pp. 59-64, March 2003.

[2] Yasushi Yagi and Naokazu Yokoya, "Omni-directional Vision: Sensors," Journal of Information Processing Society of Japan," Vol. 42, No. SIG13, pp.1-18, 2001.

[3] Yukio Kubota, "Digital Video Dokuhon," Ohme Co. 1995.

[4] Canon Co., Ltd., Sharp Co., Ltd., Sony Co., Ltd., Japan Victor Co., Ltd., "http://web.canon.jp/pressrelease/2003/hdv.html"(in Japanese), 2003

[5] Yuya Miata, Koji Hashimoto and Yoshitaka Shibata, "A New TV Conference system with Flexible Middleware for Omni-directional Camera", 8th International Workshop on Network-Based Information System(NBiS'05) pp.84-88 22 Aug. 2005.