2008－CVIM－163 (31)
2008／5／9

全方位映像のマルチキャストによる実時間ネットワークテレプレゼンスに関する研究

石川 智也 †‡　　山澤 一誠 ‡　　横矢 直和 ‡

† 産業技術総合研究所
‡ 奈良先端科学技術大学院大学

E-mail: tomoya-i@ni.aist.go.jp, yamazawa, yokoya@is.naist.jp

あらまし: 近年の計算機の高性能化やネットワークの高速化により，遠隔環境の撮影から映像提示までを実時間で実行可能な環境が整いつつあり，放送と通信の融合による次世代ネットワークメディアとして実時間ネットワークテレプレゼンスが注目されている．本研究では，遠隔地の情景を自由な視線方向でインタラクティブに観賞可能な全方位映像を用いたテレプレゼンスに焦点を当て，実時間ネットワークテレプレゼンスにおける利用者数の増加に対するスケーラビリティ及び，高臨場感なテレプレゼンスのための視線方向のみならず視点位置も自由に変更可能な映像提示を実現する．そして，これらの機能を有するハイスケーラブル自由視点テレプレゼンスシステムを提案する．

Real-Time Networked Telepresence by Multi-Casted Omni-Directional Videos

Tomoya ISHIKAWA†‡　　Kazumasa YAMAZAWA‡　　Naokazu YOKOYA‡

†National Institute of Advanced Industrial Science and Technology
‡Nara Institute of Science and Technology

Abstract: As we can capture images of a remote scene and display them in real-time, real-time networked telepresence has received much attention as a next generation network media by integrating broadcasting and network communication technologies. This study focuses on the telepresence using omni-directional videos that enables users to see a remote site interactively. The purpose of this study is to realize a high-scalable system that allows a large number of users to change not only the view-direction but also the viewpoint freely. Furthermore, we propose a novel view telepresence system that has the two functions.

1 Introduction

Telepresence is an emerging technology which provides users with a feeling of existing in remote sites by presenting the virtualized real sites [1]. As we can capture images of a remote scene and display them in real-time, real-time networked telepresence has received much attention as a next generation network media by integrating broadcasting and network communication technologies. The technology can be applied to a number of fields such as entertainments, video conferencing, medical equipments, remote surveillance, etc. Especially, live camera services which provide scenery of remote sites such as sightseeing spots are being used [2, 3]. The systems in the services employ a PTZ (Pan, Tilt, and Zoom) camera for image acquisition conventionally, and the users can change the view-direction toward an interest point by a control panel on a web browser interactively. However, the conventional systems involve latencies depending on a distance between the camera and the users for image transmission and the motion of the camera.

The telepresence system using omni-directional

cameras as a solution for the problem has been developed [4]. They can generate images in arbitrary view-directions without long latency because the images are computed from an omni-directional image. The time of the image processing does not depend on the distance and the motion of the camera. Moreover, by taking advantage of low latency, Onoe et al. [4] realize view-dependent image presentation using an HMD (Head-Mounted Display) and an electromagnetic sensor for rich presence. For the use like the live camera service, this system that generates user-centered omni-directional scenes is effective to provide the feeling of being in a remote site.

In this study, we focus on the immersive telepresence that enables users to see a remote site interactively by using omni-directional cameras for virtual tour applications. It can be presumed that a large number of people use systems of such applications. The system requires techniques of interactive image presentation for immersive telepresence. To achieve the scalability, we apply a multi-cast protocol to the telepresence system using omni-directional videos. Because omni-directional videos are independent of users' view-direction, the system only transfers same data to all of the users on the network. We propose a high-scalable telepresence system by the feature that is feasible for the multi-cast protocol. For immersive telepresence, we propose a method of novel view generation supporting omni-directional scenes and real-time processing. By applying the method to the telepresence system, the users can change not only view-directions but also viewpoints interactively. The combination of the two functions realizes a high-scalable novel view telepresence.

The rest of this paper is organized as follows. Section 2 describes the high-scalable telepresence system that achieves scalability for increase of users by the multi-cast protocol. In Section 3, we propose the high-scalable novel view telepresence system based on the system explained in the previous section. Then, Section 3 describes the experiments for considering the camera arrangement in the environment and for confirming the validity of our proposed system. Finally, conclusions and future work are given in Section 4.

2 Real-time networked telepresence by multi-casted omni-directional videos

Telepresence systems using omni-directional videos realize interactive remote viewing by transforming a part corresponding to users' view-direction from an omni-directional image which is stored or transferred in real-time into a planar perspective image [5, 4, 6, 7, 8].

Chen [5] takes an omni-directional image by stitching multiple images. On the other hand, Onoe et al. [4] use an omni-directional camera which can capture an omni-directional scene by one shot for taking omni-directional videos, and realize interactive remote viewing of dynamic environments. The users can see stored videos or live videos captured at near sites by the systems, but the systems do not support to see live remote sites via networks.

With the advent of broadband networks in recent years, real-time networked telepresence systems that enable multiple users in a site to see a remote site in each arbitrary view-direction by transferring a live video on the network have been proposed [6, 7, 8]. However, when transmitting the video on the network, they use mechanisms of server-side view generation [8] or P2P (Peer to Peer) network connection [6, 7], so that the remote viewing by multiple users in multiple sites is difficult because of limitations of the processing cost of the server and network bandwidth. Furthermore, they use some devices which are for research uses to present views, so that they are not practical.

In this study, our proposed system supports the scalability by using multi-cast protocol that enables to transfer same data to multiple clients simultaneously on networks, and excludes view-dependent processes from the server. Moreover, we improve usability of the system by web browser-based omni-directional video viewer.

2.1 Multi-cast protocol for video transmission

This section describes transmission of omni-directional videos by multi-cast protocol that enables a video server to transfer same data to multi-

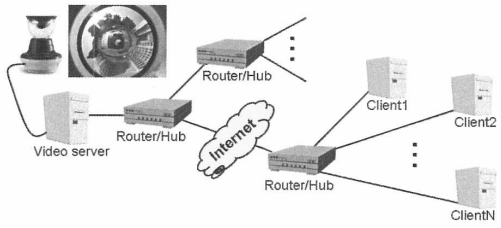

Fig. 1: Multi-casting omni-directional videos.

ple clients simultaneously. Multi-cast protocol is an extended specification of IP (Internet Protocol) for sending data to multiple targets. Recently, because of growth of broadband network and the spread of video contents, network has increased the bandwidth continuously. The multi-cast protocol has received much attention as an efficient technology of data distribution.

Networked telepresence systems have generally a server-client structure which includes a server for video distribution and some clients in the case of the application like virtual tours. When a server transfers video data that are D[Mbps] to N clients by unicast protocol, the server needs to transfer them to each client at the data rate D[Mbps]. Therefore, the server requires network bandwidth that is capable of DN[Mbps].

On the other hand, when a server transfers video data by multi-cast protocol to clients, the transferred data on network are copied by routers or switching hubs automatically and arrive in each client. Consequently the server requires network bandwidth of D[Mbps] only and so multi-cast protocol can realize efficient video distribution. But, the data arrived in each client have to be same. Telepresence systems using omni-directional videos can generate view images in arbitrary view-directions from an omni-directional image, so that they can distribute omni-directional video efficiently by multi-cast protocol. Furthermore, the cost of video distribution is independent of a number of clients.

The feature that is independent of user's view-directions is feasible for video distribution by multi-cast protocol and that can realize high-scalability which supports increase of users.

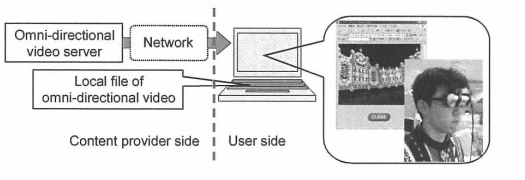

Fig. 2: Overview of real-time networked telepresence system.

2.2 Real-time networked telepresence system using multi-cast protocol

2.2.1 System overview

Our real-time networked telepresence system aims at supporting scalability for increase of users and improvement of usability. Moreover, our system supports the follows.

- Supporting live and stored omni-directional video
  Users can see not only live video but also stored video.

- View-dependent seeing and easy viewing
  When a user has an HMD with a gyro sensor, the user can see the remote scene by view-dependent presentation. When a user does not have them, the user can see contents by a viewer easily.

- Supporting various types of omni-directional cameras
  Content provider can select an omni-directional camera which is suitable for the use and content.

Figure 2 shows overview of our proposed networked telepresence system that handles omni-directional videos. The users watch the omni-directional content which is transferred from the video server in real-time via network or which is stored in users' PC by an omni-directional video viewer on a web browser. The system uses multi-cast protocol for video transmission from the server to users , so that the network traffic does not increase when the multiple users use the system at the same time. The omni-directional video viewer on a web browser generates planar perspective images by using GPU functions in real-time. By this function, the viewer enables the users to control the view-directions interactively

by mouse or keyboard operations for viewing arbitrary view-directions. Furthermore, view-dependent presentation which uses an HMD and a gyro sensor can be realized for rich presence. For supporting various types of omni-directional cameras, the system includes some correspondences between omni-directional images and a planar perspective image for planar perspective transformation in view rendering.

2.2.2 Omni-directional video contents

There are two kinds of omni-directional video contents; live video contents encoded in real-time and stored video contents encoded in advance. The live video contents are used for the purpose of providing multiple users with the same contents simultaneously like TV broadcasting. The user can see the live video contents acquired by an omni-directional camera and transferred immediately. It can be transferred to many sites by multi-cast protocol without increasing the network load. The stored video contents mainly consist of high-resolution omni-directional videos and are heavy for network. Note that stored video contents can be provided as an on-demand-service. It is difficult to transfer a high-resolution video because of the limit of standard network bandwidth.

There are several types of omni-directional cameras which consist of fish-eye lens [9], mirrors [10, 11], or a set of multiple inside-out cameras [12, 13]. Content provider has to select a camera type suitable for targets and uses. In this study, we use the omni-directional camera using a mirror; HyperOmni Vision [11]; for live video contents and a set of multiple inside-out cameras; Ladybug [13]; for stored video contents.

[HyperOmni Vision]
HyperOmni Vision consists of a hyperboloidal mirror and a common video camera and captures omni-directional images reflected by the hyperboloidal mirror (see Figure 3). By a common video camera, the acquired video can be encoded and transferred to remote sites in real-time. Therefore, this omni-directional camera suits for live video contents.

[Ladybug]
Ladybug consists of a camera unit that mounts six cameras (see Figure 4 left) and a storage unit which

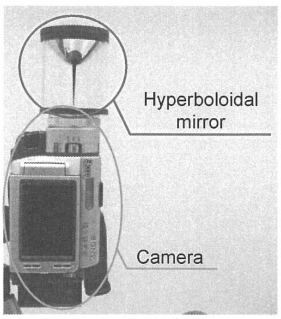

Fig. 3: HyperOmni Vision.

includes an array of HDDs (see Figure 4 right). Five cameras are placed in a horizontal ring and one camera points vertically. Each camera has XGA resolution (768×1024[pixel]) and the camera system can acquire omni-directional images covering 75% of full spherical view at 15[fps]. In the offline process, we generate the panoramic video from the acquired images using a mosaicing technique proposed by Ikeda et al. [13]. The panoramic video generated by the method has the max. 3840×1920[pixel] (see Figure 5). Ladybug has to store multiple images into a storage unit when capturing a scene , so that this camera does not suit for live video contents. But, the high-resolution videos are feasible for stored video contents.

2.2.3 Web browser-based omni-directional video viewer

A web browser is one of the most popular network applications. Especially, Internet Explorer installed on Windows PC can execute various application programs by a JAVA applet or ActiveX for providing users with interactive contents on a web page. Moreover, the JAVA applet and the ActiveX programs can be easily distributed by an automatic install function. Thus we implement an omni-directional video viewer for telepresence on a web browser.

Figure 6 shows the structure of the omni-directional video viewer. The viewer which shows an omni-directional video content to the user needs the functions of GPU (Graphics Processing Unit) and is implemented as an ActiveX program. The viewer is firstly executed by a web browser and then displays

Fig. 4: Omni-directional multi-camera system: Ladybug.

Fig. 5: High-resolution omni-directional image from Ladybug.

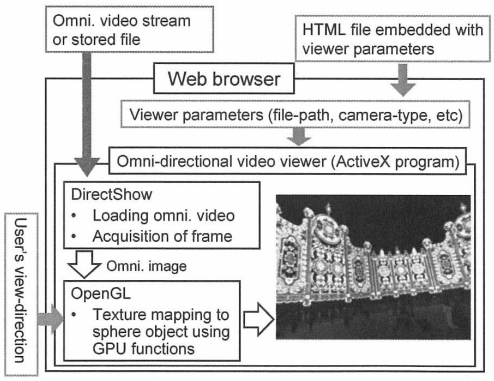

Fig. 6: Structure of omni-directional video viewer.

user's view image on a web page after transforming an omni-directional image into a planar perspective image. The details of the process are as follows.

Step1. When the user accesses the web page which provides an omni-directional video contest, the viewer is installed automatically and is executed by the web browser. At the same time, it reads the parameters embedded in the HTML file such as camera-type of an omni-directional camera, camera intrinsic parameters, file-path of content, and so on. The parameters are embedded by a content provider, and so the user does not need to take care of them.

Step2. The viewer accesses the omni-directional video stream or the stored file using DirectShow functions. DirectShow functions support the various video formats such as AVI, MPEG, WMV, and so on, and can acquire each frame of the video stream.

Step3. After the acquired frame is transferred to a texture memory on GPU, the viewer transforms an omni-directional image into a planar perspective image by OpenGL functions with GPU according to the user's view parameters are taken from a mouse, a keyboard, or a gyro sensor. In this process, hardware texture mapping to a sphere object is used for the transformation in real-time by the technique proposed by Onoe et al. [4].

Step4. Finally, the viewer draws the view image to the window of the web browser. These functions are synchronized with the frame rate of inputted video. The user can look around the omni-directional video by using a mouse-drag operation interactively. When the user uses an HMD with a gyro sensor, the viewer draws the video on full screen by setting the "full screen" parameter wrote in the HTML file.

2.3 Experiments

We have implemented real-time networked telepresence system described above and have carried out the experiment using the live video content to verify the functions. In this experiment presumed to apply the system to applications like TV broadcasting, the car mounted with an omni-directional camera distributes the omni-directional live video by multi-cast protocol directly. We confirm that multiple users can see the live video interactively and simultaneously.

Table 1: Specifications of car for video acquisition.

Omni. camera	SONY DCR-TRV900 + Hyperboloidal mirror
PC for capturing, encoding , and multi-cast distribution	Pentium4 2.53[GHz] Memory 1[GB] WindowsXP
Wireless network	IEEE802.11g
Car	Nissan ELGRAND

2.3.1 System configuration

Figure 7 illustrates the configuration of the real-time networked telepresence system in the experiment using live video contents. The system consists of the car mounted with HyperOmni Vision, a PC for video capturing, encoding, and multi-cast distribution, the omni-directional video viewer on the PCs for remote viewing, and wired/wireless network.

The car captures an omni-directional live video (resolution: 720×480[pixel], frame rate: 15[fps]) as it drives in our campus. The captured live video is transferred to the PC for video capturing, encoding, and multi-cast distribution through i.Link. Then, the PC as a live video server encodes the omni-directional video to Windows Media Format (data rate: 832[kbps]) by the Windows Media Encoder [14]. The encoded video (resolution: 640×480[pixel], frame rate: 24[fps]) is distributed to the wireless network by using multi-cast protocol. Table 1 shows the specifications of the car for video acquisition.

The omni-directional video viewer receives the distributed video and decodes it. Then, the viewer transforms the decoded omni-directional image into the planar perspective image. The users can see the distributed live video interactively by the viewer on a PC.

2.3.2 Telepresence using multi-casted live videos

In this experiment, the five PCs receive the omni-directional live video simultaneously. We have confirmed that the users of the PCs can see the live video in arbitrary view-directions. Examples of the screenshots of the omni-directional viewers are shown in Figure 8. The view images displayed on the web

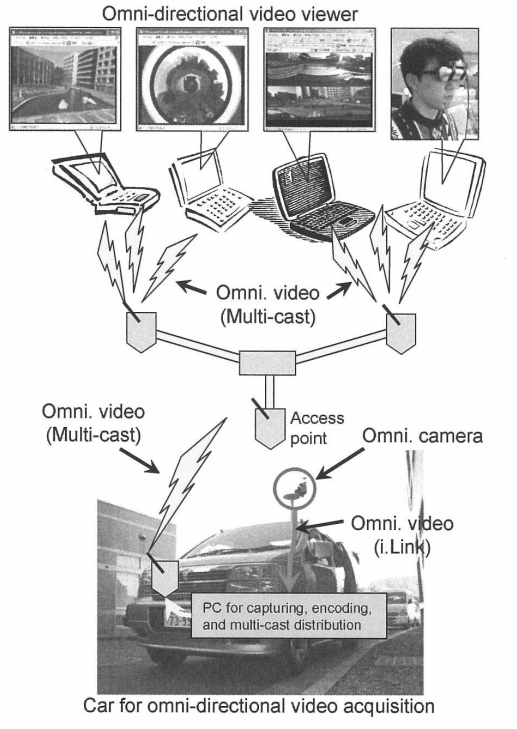

Fig. 7: System configuration in the experiment using live video contents.

browser are renewed at 24[fps] that is same as the distributed video and the delay between capturing an image by the omni-directional camera and displaying the image is about 10[s]. The major cause of the delay is buffering for encoding and decoding of the omni-directional video. We have confirmed that the network load is around 832[kbps] constantly when increase of the PCs receiving the live video.

For rich presence, one of the PCs for remote viewing employs an HMD with a gyro sensor (INTERSENSE Inc. InterTrax2) that measures user's view-direction and carries out view-dependent presentation (see Figure 9). The user can change the view-direction by own motion and see the scenery like surfing on the car. In the case using the gyro sensor, the delay between changing user's view-direction and displaying the view image corresponding to the view-direction is about 25[ms]. Wireless networks are generally more unstable than wired them , so that viewing the uni-casted live video by multiple users is

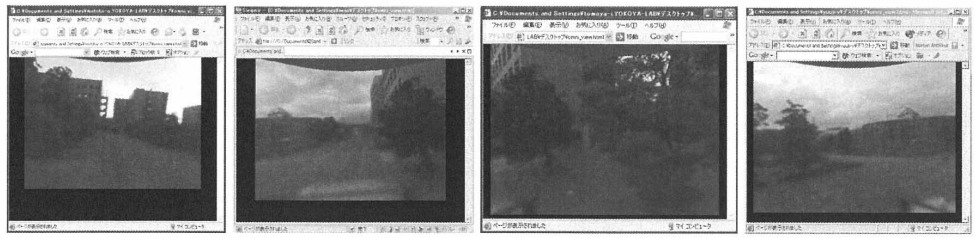

Fig. 8: Examples of screenshots of omni-directional video viewers.

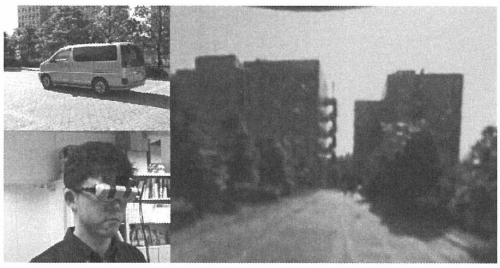

Fig. 9: View-dependent presentation using HMD with gyro sensor (left top: car mounted with omni. camera, left bottom: appearance of user, right: user's view image).

difficult. But, by multi-casting of the live video, our proposed system allows the multiple users to view the video stably.

2.4 Summary

In this section, we have proposed high-scalable real-time networked telepresence that can support simultaneous uses of multiple users by multi-casting omni-directional videos. By the omni-directional video viewer considering usability, the system enables the users can see omni-directional video contents easily. Moreover, the system can support not only live video distributed in real-time like TV broadcasting but also stored video used for on-demand services.

In the experiment of live video contents, we have confirmed that the proposed system is scalable to increase of users. The displayed video becomes rough because the resolution of the source video captured by the standard definition camera. In future, we will come to be able to use high-resolution videos for live video contents by the advent of high-performance PCs that can encode the video from Ladybug or HyperOmni Vision using an HD camera in real-time.

In the current implementation, the delay between capturing and displaying images is about 10[s]. The major cause is buffering for encoding and decoding the video, so that the system should employ a codec of videos with low delay when the system is applied to communication use between remote users.

3 Novel view telepresence using multiple omni-directional videos

This section describes high-scalable novel view telepresence that allows multiple users to see a virtualized remote site in arbitrary view-directions at arbitrary viewpoints interactively based on real-time networked telepresence system described in previous section by infusing a method of novel view generation. The telepresence systems introduced in previous section take a degree of freedom about view-directions by using omni-directional videos. However, in the systems, the users' viewpoints are restricted onto the camera-path, which is same as shooting camera positions , so that the systems can not present view images corresponding to users' motion completely.

For realizing the degree of freedom of users' viewpoints, the various techniques of novel view generation have been proposed. Koyama et al. [15] have proposed the live telepresence system for viewing soccer scenes. They also proposed a method for novel view generation based on the assumption that the viewpoints are far from soccer players. Because of the assumption, the method approximately generates novel view images using a billboarding technique in online processes. Their system enables multiple users

to see soccer scenes at arbitrary viewpoints interactively via internet by transferring data needed by the view generation only. Accordingly, the system has to manage each user and the function induces heavy network load and computational costs when a large number of users access to the system.

In this section, we first propose the novel view generation method for realizing a high-scalable telepresence system. The method virtualizes the whole shooting scene which includes both inside-out and outside-in observations of camera positions simultaneously from multiple omni-directional videos in real-time. Then, we propose the high-scalable novel view telepresence system by applying the proposed method. The system places omni-directional cameras at different positions and the captured multiple videos are distributed by multi-cast protocol from the live video server. The multiple clients generate novel view images from distributed videos according to users' view-direction and viewpoint taken from position and posture sensors in real-time and present the generated views to the users with a feeling of walking in a remote site. Finally, we demonstrate our proposed system and carry out the subjective evaluation for confirming the validity.

3.1 Novel view generation from multiple omni-directional videos

In this section, we describe the novel view generation method from multiple omni-directional videos. The an omni-directional novel view image from a set of omni-directional images which are independent of view-directions and viewpoints. Our method segments input images into static and dynamic regions and feasible techniques are applied to each region. In addition, for reducing computational cost, the method estimates existing regions of dynamic regions from multiple silhouettes on input images.

The method of novel view generation carries out following procedure (see Figure 10). Firstly, omni-directional images are acquired by HyperOmni Visions located in an environment for virtualzing whole shooting scene. Secondly, the acquired images are segmented into static and dynamic regions by using the robust background subtraction technique [16]. Thirdly, a novel view image of the static region is generated by a morphing-based technique, and a novel view image of the dynamic region is generated by computing visual hulls. Finally, the two generated images are synthesized as a complete novel view image. In Sections 3.1.1 and 3.1.2, we explain the outlines of novel view generations for static and dynamic regions, respectively.

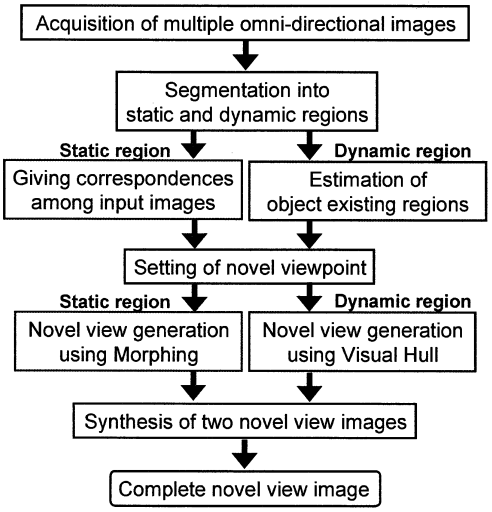

Fig. 10: Flowchart of novel view generation.

3.1.1 Morphing-based rendering for static regions

A morphing-based rendering technique is used to generate an omni-directional novel view image of static regions from omni-directional images. This technique needs 2D correspondences among input images. Note that the corresponding points are given in advance. The processes of the image generation based on morphing is as follows.

Step1. The 3D positions of corresponding points are computed by omni-directional stereo from given 2D corresponding points (see Figure 11).

Step2. The 3D points computed by Step1 are projected onto the novel view image plane.

Step3. Triangulated patches in the omni-directional novel view image are generated based on the projected points using Delaunay's triangulation.

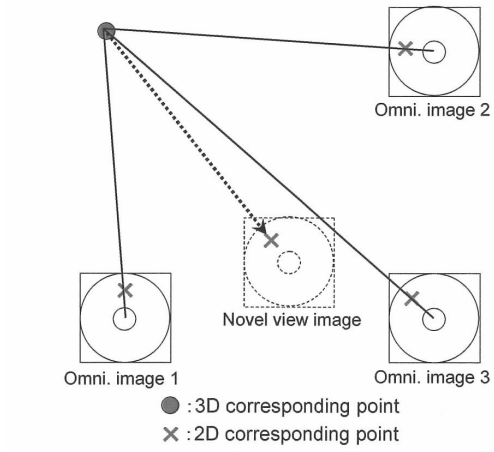

Fig. 11: Estimation of 3D corresponding point from 2D corresponding points.

Step4. The omni-directional novel view image is rendered by transforming and blending the parts of input omni-directional images which correspond to the triangle patches.

In this method, the calculation for pixels within triangles and the blending are executed by using OpenGL functions with GPU. Therefore, we can generate a novel view image without high CPU cost. For more details, see the reference [17].

3.1.2 Visual hull-based rendering for dynamic regions

The novel view image of dynamic regions is generated by computing visual hulls. Visual hull constructed by the intersection of silhouette cones from images of several viewpoints defines an approximate geometric representation of an object. In general, the shape-from-silhouette technique [18] is used for computing the visual hull. The method employs the voxel representation of a space. Therefore, the cost of computing the visual hull and the size of data become huge for a wide area. Our proposed method uses the Image-based Visual Hull technique [19] for computing visual hulls with reduced cost. The technique does not use voxels but generates a novel view image by estimating the penetration of visual hull by the ray from the novel viewpoint. The overview of

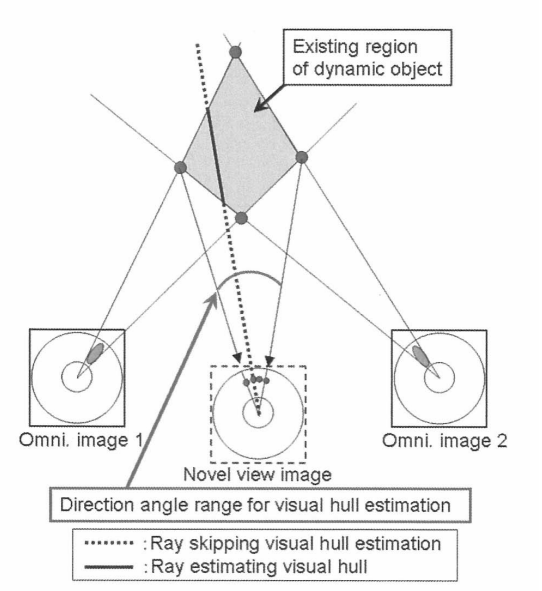

Fig. 12: Reduction of computational cost by estimation of existing region of dynamic objects.

the process is as follows.

Step1. For each pixel in the novel view image, a ray connecting a pixel and the novel viewpoint is projected onto input omni-directional images.

Step2. On the line of the ray, we search for a segment in which all of the projected lines intersect with dynamic regions. If it is found, the ray is judged to penetrate the visual hull. If it is not found, the ray does not penetrate the visual hull.

Step3. If the ray penetrates the visual hull, the pixel in novel view image is colored. The point on the intersection segment nearest to the novel viewpoint is projected onto the input image that is selected by the similarity of the angle formed by the novel viewpoint, the intersection point, and the viewpoint of input image. The color of the projected point is decided as the color of the pixel in the novel view.

Step4. The processes consisting of Step1-3 are executed for all the pixels in the novel view image.

Processing for all the pixels in the novel view image needs high computational cost. For reducing the

cost, we have to limit the pixels that need to estimate visual hulls by computing object existing regions re-projected on novel view images. The object existing regions are represented as sets of points crossing lines which mean borders of an angle range detecting an object silhouette on an input image (see Figure 12). See the reference [20] for the details of the estimation for object existing regions. Our method needs to estimate the visual hull only on the limited pixels which are included in the angle ranges of the object existing regions on the novel view image. Furthermore, on a ray estimating visual hull, we skip the estimation when the ray passes through outside of object existing regions.

3.2 High-scalable novel view telepresence system

3.2.1 System overview

This section describes high-scalable novel view telepresence system that can present novel views to multiple users simultaneously in real-time using the method of novel view generation described in Section 3.1. The system distributes omni-directional live videos captured at remote sites by multi-cast protocol, and the users see the remote site with walking and viewing like being the remote site immersively.

The top of Figure 13 illustrates the overview of the proposed system. The system consists of the live video server for distributing omni-directional live videos and novel view generators for generating novel view images from the distributed videos. The live video server realizes the scalability by multi-casting the videos from server-connected cameras. The novel view generators present view-dependent images to the users equipped with an HMD and a position and posture sensor. To the users using a mouse and a keyboard, the generators display novel view images onto an LCD, and the users control the viewpoints and view-directions by mouse and keyboard operations.

3.2.2 Processes of server and generator

Both of the live video server and the novel view generators carry out generate pre-computable data in offline for efficient online processing. In addition, the server computes the background subtraction independent of users' viewpoints and view-direction for distributed computing. In the below, the online and offline processes are explained following the flowchart of Figure 13 bottom.

(1) Offline process
[Live video server]

- Camera intrinsic and extrinsic parameters located at some positions are calibrated using the method proposed by Negishi et al. [21].
- Regions of corners and projected camera on input images are not needed to process. The mask images are generated for invalidating the regions.
- An initial background images are captured for the background subtraction.

[Novel view generator]

- 2D corresponding points among input images are given manually in advance.
- 3D ray vectors corresponding to each pixel connected to the pixel and a novel viewpoint are needed for computing visual hulls so that the ray vectors are pre-computed.
- As well as the live video server, mask images for invalidating the regions not needed to process are generated.

(2) Online process
[Live video server]

1. After capturing omni-directional images, mask images (1bit/pixel) that are results of the background subtraction are generated.
2. The captured images are encoded into JPEG.
3. The encoded images and mask images are annexed to a time code meaning the time of image capturing. They are multi-casted to network after packetizing.

[Novel view generator]

1. The multi-casted images are received and their time codes are checked. When all of the time codes are coincident, following processes are executed.

2. The encoded images are decoded.

3. Users' viewpoint and view-direction are measured by a position and posture sensor or mouse and keyboard.

4. According to the user's view information, a novel view image are generated. When computing visual hulls, in addition to the reducing computational costs based on estimation of object existing regions, the pixels toward user's view-direction are estimated.

5. After the novel view generation, The generated image on the frame buffer of GPU is used as a texture for fast transforming the omni-directional image into a planar perspective image using Onoe's method [4].

6. The transformed image is displayed onto an HMD or an LCD.

The system carries out the above processes iteratively and realizes the novel view telepresence.

3.3 Experiments

3.3.1 System configuration

We have implemented the prototype system of the high-scalable novel view telepresence illustrated in Figure 13. The details of the live video server and the novel view generators are given in the following.

[Live video server]

The live video server is connected with three omni-directional cameras, HyperOmni Visions (Suekage Inc., SOIOS 55-Cam) by IEEE1394 and they are located in a remote site. The server and the novel view generators are connected to Gigabit Ethernet. The server distributes the omni-directional images from the cameras and mask images as results of the background subtraction to the novel view generators by multi-cast protocol.

[Novel view generators]

The novel view generators receive the omni-directional images and binary mask images and generate novel view images. When generating the novel view images, the electromagnetic sensor (Polhemus Inc., 3SPACE FASTRAK) or mouse / keyboard measure the users' view-direction and viewpoints. For measuring wide range compared with a standard transmitter, the system uses the long ranger (Figure 14 right). The user equips an HMD (Olympus Inc., FMD-700) with a receiver of the electromagnetic sensor (Figure 14 left) and sees the remote scenery by walking like being there.

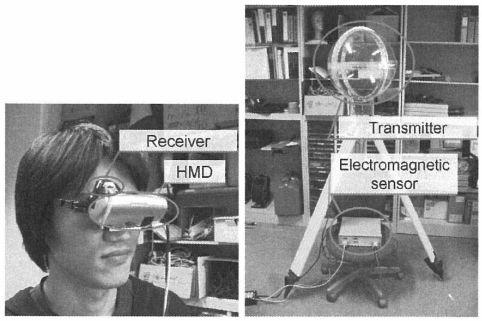

Fig. 14: User-worm HMD with sensor (left) and electromagnetic sensor (right).

Table 2: Specifications of omni. camera, live video server, and novel view generators

Omni. camera	Resolution: 640x480[pixel] Max. frame rate: 15[fps] Field of view: Horizontal:360[deg] Vertical:62[deg]
Live video server	CPU: Intel Pentium4 3.2[GHz]
Novel view generator1	CPU: Intel PentiumD 940 GPU: nVidia GeForce7300GS
Novel view generator2	CPU: Intel Pentium4 3.2[GHz] GPU: nVidia GeForce6600GT

Table 2 shows the specifications of the omni-directional cameras, the live video server, and the novel view generators.

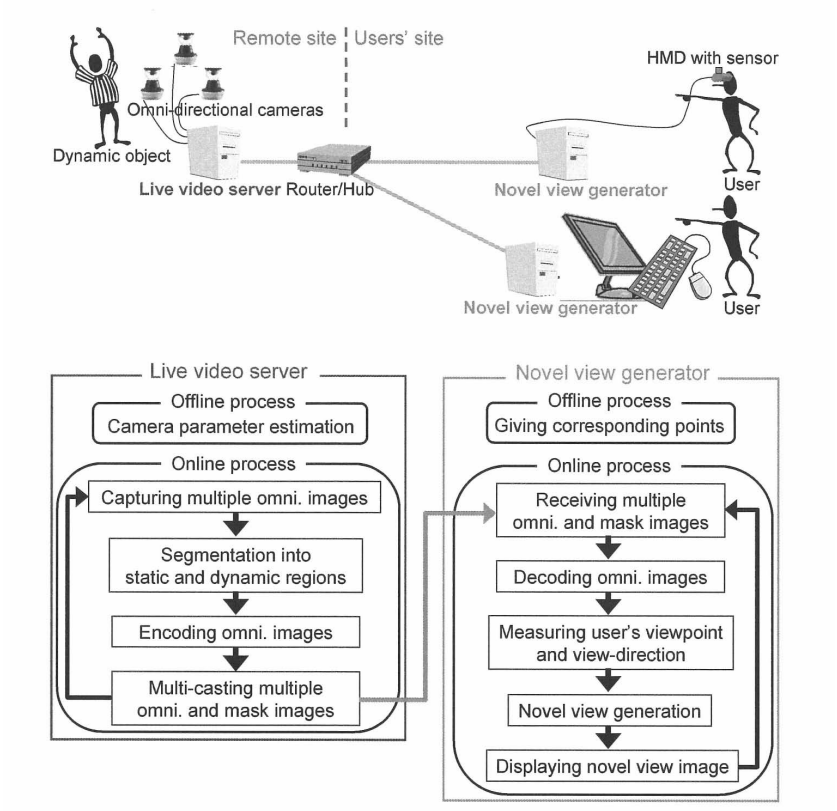

Fig. 13: System overview (top) and flowchart of system (bottom).

3.3.2 Functional demonstration

We carried out the functional demonstration of the high-scalable novel view telepresence by using the prototype system. In this demonstration, we confirm that the proposed system enables multiple users to see the remote site at arbitrary viewpoints in arbitrary view-directions interactively without increase of network load.

One of the two users (user1) is presented view-dependent images and the other user (user2) sees novel view images on an LCD by changing the viewpoints and view-directions using a mouse and a keyboard. The omni-directional cameras are positioned forming a triangle (see Figure 15 top) which spaces about 2[m] between cameras. The estimated camera positions and postures are shown in Figure 15 bottom. The black points and three pyramids mean the calibration markers and the cameras. The 70 corresponding points are manually established by mouse clicking only for the static scene. The live video server distributes data at 10[fps] by the limitation of the capability.

Figure 16 and 17 show the appearances of both remote and user sites and the generated novel view images. The remote site in Figure 16 does not include dynamic objects. On the other hand, the scene of Figure 17 includes two dynamic objects. We can confirm that the both users can see the dynamic remote site interactively at the same time. Each frame of novel view images presented to each user is generated in around 200[ms]. The details of the time are shown in Table 3. However, the time for the novel view generation of dynamic regions is not constant because the time depends on the region size of dynamic objects re-projected on the novel view image.

The network load from the server is shown in Fig-

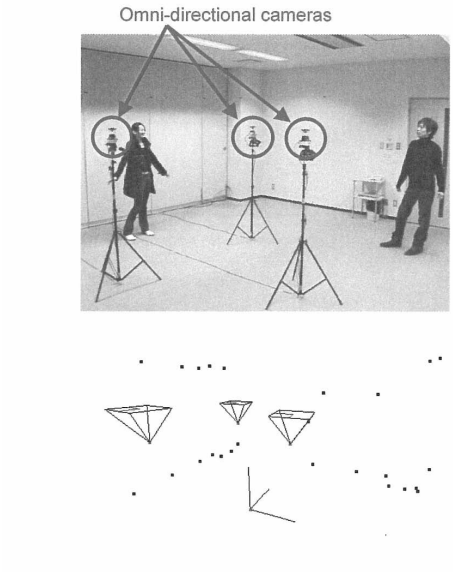

Fig. 15: Omni-directional cameras located in remote site (top) and estimated camera positions (bottom).

Table 3: The details of processing times in novel view generator1.

Receiving omni. & mask images	about 50[ms]
Decoding omni. images	about 30[ms]
Measuring view info.	about 8[ms]
Morphing-based rendering	about 20[ms]
Visual hull-based rendering	about 100[ms]
Perspective transformation	about 1[ms]

ure 18. The server started to distribute data at time 0 and then, 6 seconds later, user1 started telepresence. User2 joined in the system at 24 seconds after starting the server. In spite of increasing the number of users, the network load is constant.

3.3.3 Subjective evaluation

The section describes the subjective evaluation of validity of the novel view telepresence system by subjects. Our proposed system has realized high-scalability for a number of users and interactivity for changing viewpoints compared with the system (Section 2) using omni-directional videos. However,

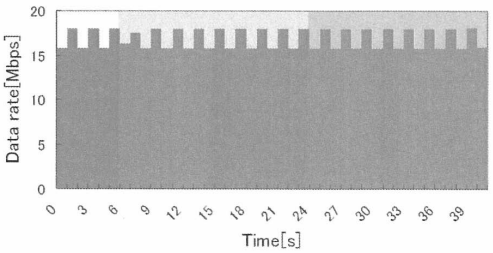

Fig. 18: Network load after starting live video server.

by the novel view rendering, the quality of view images presented to users becomes low. The low quality images may prevent presence of view-dependent image presentation. With that, we evaluate the validity by comparison between our proposed system and the system that respects the image quality rather than view-correspondence between user's and novel viewpoint. In the evaluation, the subjects judge that "which is the system good about image quality, view-correspondence, and presence". The rest of this section gives the conditions of the evaluation and the result.

[Conditions of evaluation]

We first explain about system A, B for this evaluation. System A which aspects the image quality presents planer perspective images transformed from input omni-directional images to the subjects without novel view generation. In addition, the images of a camera nearest to a subject's viewpoint are used as the input images. System B is just our proposed novel view telepresence system. The configurations of both the systems are the same as the prototype system described above (Section 3.3.1). But, in this evaluation, we use stored multiple omni-directional videos for unification of a virtualized scene (see Figure 19) presented to each subject. The virtualized scene includes dynamic objects like above demonstration.

After the subjects experience both the systems for a minute each, they select "A is better than B.", "B is better than A.", or "A and B are equal." about "image quality", "view-correspondence", and "presence". We explained that "image quality" is better as you feel natural or beautiful, "view-correspondence" is better as the presented images change according to your motion, and "presence" is better as you feel

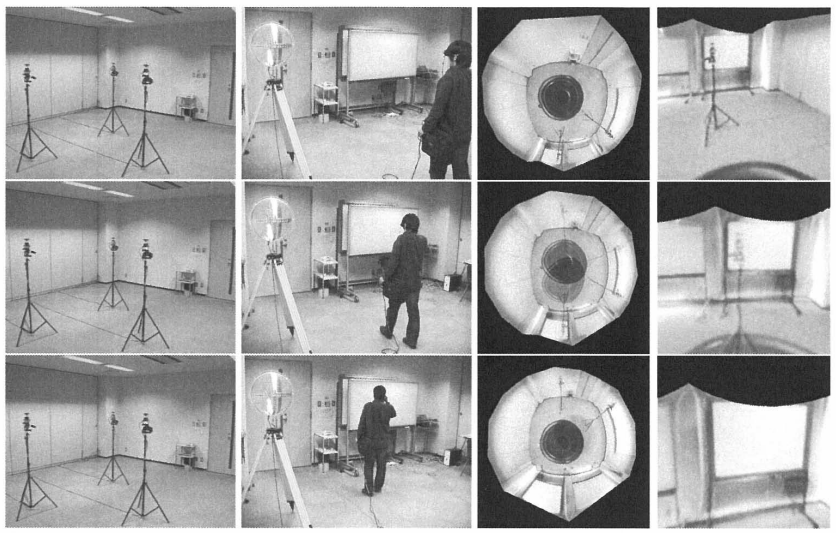

Fig. 16: Appearance of remote and user's sites (a, b), generated novel view images (c), and planar perspective images (d) in the demonstration without dynamic objects.

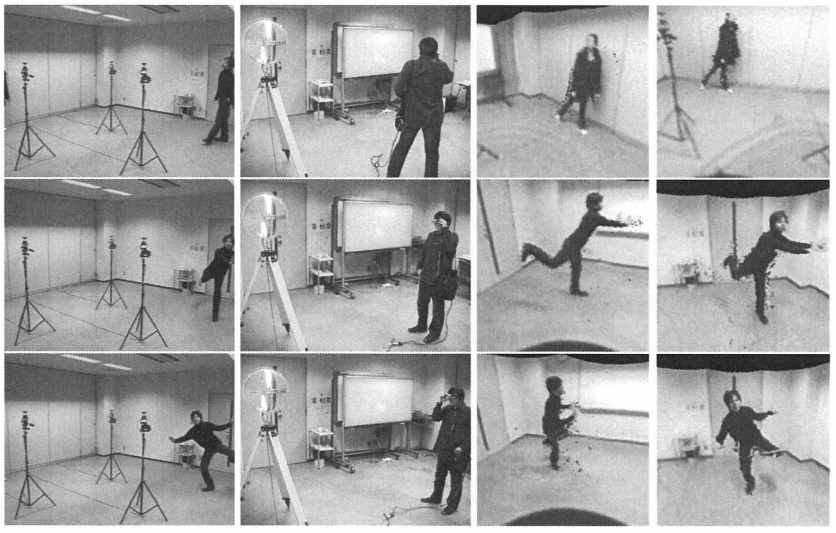

Fig. 17: Appearance of remote and user's sites (a, b) and view images presented to user1 and 2 (c, d) in the demonstration with dynamic objects.

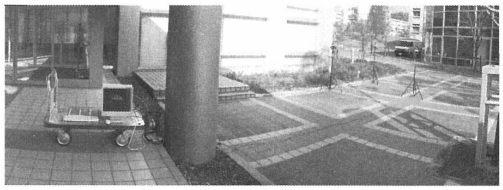

Fig. 19: Environment presented to subjects in subjective evaluation.

being in the presented scene. Moreover, we made the subjects walking as wide as possible. The subjects experience two cases about image presentation in each system. One of the two cases was the same frame rate of view images between the systems because each system has different frame rate by different processes. The other case was not the same frame rate.

[Result of evaluation]

Table 4 shows the result of the evaluation by twelve subjects. The colored row means that the evaluation item on the case has statistical difference among "A is better than B.", "B is better than A.", and "A and B are equal." at the 0.05 level of significance.

About image quality, system A was evaluated as better in both the cases. From this result, we can confirm that the most of subjects understood the view images generated by system B are under image quality compared with the input images significantly.

In contrast, as to view-correspondence, system B was evaluated as better in both the cases. The result is considered that system B made the subjects to understand changing the viewpoints continuously compared with system A that switches over the input images selectively.

About presence, system B was judged to better one in the case of the same frame rate. System A does not generate novel view images, so that the frame rate of view images becomes higher than system B. Thus, in the case of the different frame rate, they were no odds. But, we consider that the subjects felt that system B that made the subjects to understand the change of viewpoints was better on the case of the same frame rate.

Table 4: Result of evaluation by twelve subjects.

		A>B	A=B	A<B
Same frame rate	Image quality	11	0	1
	View-correspondence	1	1	10
	Presence	5	4	3
Different frame rate	Image quality	11	0	1
	View-correspondence	1	1	10
	Presence	2	1	9

3.4 Summary

In this section, we have proposed high-scalable novel view telepresence system based on the real-time networked telepresence system proposed in the previous section by using the method of novel view generation. We have confirmed that the proposed system has the scalability and can present view-dependent images to the users. In addition, we have confirmed that the novel view images are low quality compared with the input images and the effect of view-dependent image presentation. Especially, when the frame rates are the same, the proposed system is better than the system of the fixed viewpoints. In near future, the novel view images will be generated at video rate by advent of high-performance PCs.

The high-scalable novel view telepresence system proposed in this section uses JPEG encoding for image compression, so that the users can see a remote site without long delay. Because of that, the system can be applied to not only the applications of virtual tours but also a field of remote communications like rich video conferencing. Otherwise, applying the system to a child facility, a large number of parents can see the appearance of their children with a feeling of being there. But, in the case of that, we have to improve the interface which can be used by anyone easily to view the remote site.

4 Conclusion

In this study, we have proposed the high-scalable novel view telepresence that enables multiple users to see a remote site at arbitrary viewpoints in arbitrary view-directions interactively for virtual tour applications. For realizing the system, we support the scalability for a number of users and the novel

view generation in real-time. The conventional systems transfer the video on network by uni-cast protocol, so that network bandwidth and computational costs are required relatively to a number of users. By contrast, we realize high-scalable telepresence using multi-cast of the omni-directional videos by taking advantage of view-independence of omni-directional videos without increase of the bandwidth and computational costs.

The rest of the problem, the view-dependent presentation according to users' viewpoint is supported by novel view generation from multiple omnidirectional videos in real-time. Our method segments input images into static and dynamic regions and applies feasible techniques to each region. In addition, for reducing computational cost, the method estimates existing regions of dynamic objects from multiple silhouettes on input images. By the strategies, we realize the virtualization for omni-directional cenery and real-time processing. The high-scalable novel view telepresence system has been realized by integrating the method into the high-scalable telepresence system.

The future work includes realization of a framework that can handle various types of cameras uniformly and a browser that does not make the users to take care of system configurations such as camera positions.

References

[1] S. Moezzi. Immersive telepresence. *IEEE MultiMedia*, Vol. 4, No. 1, pp. 17–56, 1997.

[2] EarthCam, Inc. Earthcam - where the world watches the world. *http://www.earthcam.com/*, 1996.

[3] Paris-Live.com. Eiffel tower webcams -paris france video-news-. *http://www.paris-live.com/*, 2000.

[4] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya. Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding (CVIU)*, Vol. 71, No. 2, pp. 154–165, 1998.

[5] S. E. Chen. QuickTime VR - an image-based approach to virtual environment navigation. *In Proc. of SIGGRAPH'95*, pp. 29–38, 1995.

[6] T. Yamamoto and M. Doi. PanoVi: Panoramic movie system for real-time network transmission. *In Proc. of IEEE 4th Workshop on Multimedia Signal Processing*, pp. 389–394, 2001.

[7] S. Morita, K. Yamazawa, and N. Yokoya. Internet telepresence by real-time view-dependent image generation with omnidirectional video camera. *In Proc. of SPIE Electronic Imaging*, Vol. 5018, pp. 51–60, 2003.

[8] ViewPLUS Inc. LiveSPHERE. *http://www.viewplus.co.jp/products/ livesphere/index.html*, 2003.

[9] Z. L. Cao, S. J. Oh, and E. L. Hall. Dynamic omnidirectional vision for mobile robots. *In Proc. of SPIE Intelligent Robots and Computer Vision*, Vol. 579, pp. 405–414, 1985.

[10] Y. Yagi, Y. Nishizawa, and M. Yachida. Estimating location and avoiding collision against unknown obstacle for the mobile robot using omnidirectional image sensor copic. *In Proc. of Int. Workshop on Intelligent Robots and System*, pp. 909–914, 1991.

[11] K. Yamazawa, Y. Yagi, and M. Yachida. New real-time omni-directional image sensor with hyperboloidal mirror. *In Proc. of 8th Scandinavian Conf. on Image Analysis*, Vol. 2, pp. 1381–1387, 1993.

[12] K. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura, and N. Yokoya. Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. In Proc. of 14th IAPR Int. Conf. on Pattern Recognition (ICPR'98) , pp. 485–489, 1998.

[13] S. Ikeda, T. Sato, M. Kanbara, and N. Yokoya. Panoramic movie generation using an omnidirectional multi-camera system for telepresence. *In Proc. 13th Scandinavian Conf. on Image Analysis (SCIA2003)*, pp. 1074–1081, 2003.

[14] Microsoft Corporation. Windows media encoder 9 series. *http://www.microsoft.com/windows/windowsmedia /9series/encoder/default.aspx*, 2002.

[15] T. Koyama, I. Kitahara, and Y. Ohta. Live mixed-reality 3D video in soccer stadium. *In Proc. of 2nd IEEE/ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR03)*, pp. 987–990, 2003.

[16] K. Yamazawa and N. Yokoya. Detecting moving objects from omni-directional dynamic images based on adaptive background. *In Proc. of 10th IEEE Int. Conf. on Image Processing (ICIP2003)*, Vol. III, pp. 953–956, 2003.

[17] K. Tomite, K. Yamazawa, and N. Yokoya. Arbitrary viewpoint rendering from multiple omnidirectional images. *In Proc. of 16th IAPR Int. Conf. on Pattern Recognition (ICPR2002)*, pp. 987–990, 2002.

[18] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 16, No. 2, pp. 150–162, 1994.

[19] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. *In Proc. of SIGGRAPH2000*, pp. 369–374, 2000.

[20] S. Morita, K. Yamazawa, and N. Yokoya. Networked video surveillance using multiple omnidirectional cameras. *In Proc. of IEEE Int. Sympo. on Computational Intelligence in Robotics and Automation*, pp. 1245–1250, 2003.

[21] Y. Negishi, J. Miura, and Y. Shirai. Calibration of omnidirectional stereo for mobile robots. *In Proc. of 2003 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2600–2605, 2004.