Internet Telepresence by Real-Time View-Dependent Image Generation with Omnidirectional Video Camera

Shinji Morita, Kazumasa Yamazawa and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0101, Japan

ABSTRACT

This paper describes a new networked telepresence system which realizes virtual tours into a visualized dynamic real world without significant time delay. Our system is realized by the following three steps: (1) video-rate omnidirectional image acquisition, (2) transportation of an omnidirectional video stream via internet, and (3) real-time view-dependent perspective image generation from the omnidirectional video stream. Our system is applicable to real-time telepresence in the situation where the real world to be seen is far from an observation site, because the time delay from the change of user's viewing direction to the change of displayed image is small and does not depend on the actual distance between both sites. Moreover, multiple users can look around from a single viewpoint in a visualized dynamic real world in different directions at the same time. In experiments, we have proved that the proposed system is useful for internet telepresence.

Keywords: Internet Telepresence, Digital Video Transport System, Omnidirectional Imaging, HyperOmni Vision, Virtual Reality

1. INTRODUCTION

There are many requirements that a dynamic real world should be displayed to a remote user with rich presence in telerobotics, video conference, virtual tour, and so on. Virtual reality technology which aims at presenting a user a scene of a remote site and providing the user with the natural feeling of existing in the distance is called telepresence or immersive telepresence.

In most conventional telepresence systems, fixed video cameras are generally used in order to acquire a scene of an observed world. However, the visual field of an observer is usually restricted by the angle of field of a lens in this method. Accordingly, it is indispensable to acquire a wide range of a scene without limiting the angle of view of a camera in order to realize better communication. For the above problem, active cameras such as panning and tilting cameras which are placed on a rotation stand are utilized to impressively present a user view-dependent images of a scene in the distance.[1–5] This often suffers from the time delay from the change of user's viewing direction to the change of displayed image. The time delay is mainly caused by both the communication between a user site and an observed world and the control of cameras so as to follow the user's viewing direction. Especially, the former factor depends on the actual distance between the user and remote sites. Moreover, multiple users can not look around a visualized dynamic real world in different directions at the same time. For the above problems, we proposed some telepresence systems using an omnidirectional video camera.[6] Those systems acquire omnidirectional images of a dynamic real world using an omnidirectional video camera, and then generate view-dependent images displayed to multiple users in real-time. In this method, the time delay from the change of user's viewing direction to the change of displayed image is small and does not depend on the actual distance between both sites. However, those systems are implemented without the consideration of the internet, that is to say, omnidirectional images are not transported from the observed site to the user site through the internet; that is, the systems were not networked.

Further author information: (Send correspondence to Shinji Morita)
Shinji Morita: E-mail: shinji-m@is.aist-nara.ac.jp, Telephone: +81-743-72-5296; Fax: +81-743-72-5299
Kazumasa Yamazawa: E-mail: yamazawa@is.aist-nara.ac.jp, Telephone: +81-743-72-5291
Naokazu Yokoya: E-mail: yokoya@is.aist-nara.ac.jp, Telephone: +81-743-72-5290

We have developed a new telepresence system which works on internet. Our system transports omnidirectional video streams through the internet, besides the time delay from the change of user's viewing direction to the change of displayed image does not depend on the actual distance between both sites. Therefore, the present system is applicable to real-time telepresence in the situation where the real world to be seen is far from an observation site. Moreover, multiple users can look around from a single viewpoint in a visualized dynamic real world in different directions at the same time via internet.

This paper is structured as follows. In Section 2, the outline and each process of internet telepresence system are described. We describe some experimental results, as well as some prototype of telepresence system implemented with some kinds of network in Section 3. We discuss them in Section 4. Finally, Section 5 concludes the paper.

2. INTERNET TELEPRESENCE SYSTEM

In this section, we describe the outline and each process of internet telepresence system.

2.1. Outline of the System

The proposed internet telepresence system is realized by the following three steps:

1. video-rate omnidirectional image acquisition,
2. omnidirectional image transportation through the internet,
3. real-time view-dependent perspective image generation from an omnidirectional video stream.

Our system is a peer to peer system, in which a computer that is placed in an omnidirectional video camera side is called the sender and a computer that is placed in an observation side is called the receiver. Omnidirectional images are transported from the sender to receiver over IP4 or IPv6. Moreover, in the receiver side, multiple users can individually look their own view-dependent perspective images.

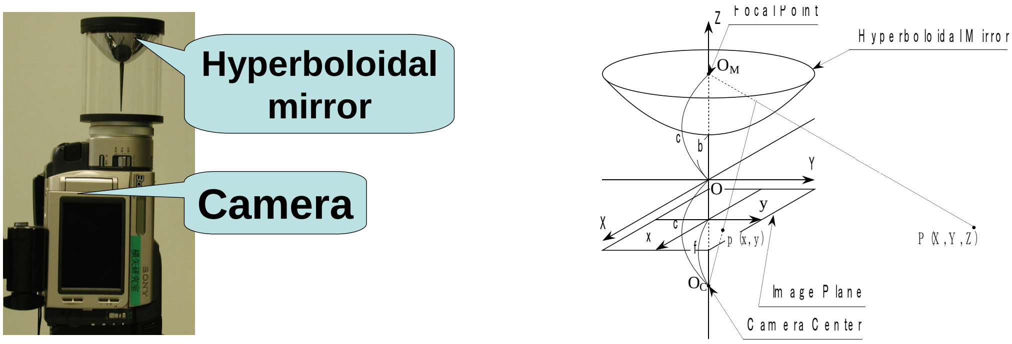
2.2. Acquisition of Omnidirectional Images

Video-rate acquisition of an omnidirectional image of a dynamic real scene requires specialized cameras. A variety of sensors have been developed to acquire omnidirectional visual information on a 3D environment.[7,8] Optics of such sensors use a fish-eye lens,[9] a conic mirror,[10] a spherical mirror,[11] a hyperboloidal mirror,[12] or a paraboloidal mirror.[5] We employ the HyperOmni Vision,[12,13] which is a catadioptric camera using a hyperboloidal mirror. HyperOmni Vision satisfies the following two requirements in the telepresence.

1. video-rate imaging of a dynamic environment.
2. satisfaction of perspective optics (single view-point constraint).

In the following we briefly describe the geometry and characteristics of the camera. Figure 1 shows an appearance of HyperOmni Vision and its optical layout. HyperOmni Vision is mainly composed of a video camera and a hyperboloidal mirror as illustrated in Figure 1(a). Therefore, as shown in Figure 2, the camera can capture a 360-degree circular omnidirectional image of a 3D environment around the sensor. The hyperboloidal mirror has two focal points O_M and O_C as shown in Figure 1(b) and all rays coming into O_M are reflected by the mirror surface and pass through O_C. The projection center of a camera lens is placed at O_C. That is, HyperOmni Vision satisfies the single view-point constraint and thus the omnidirectional image captured by the camera can be converted to the perspective image (Figure 3) and the panoramic image seen from O_M. This feature is important for virtual reality and telepresence applications. Other omnidirectional sensors using conic or spherical mirrors do not have such a good feature. This is the main reason why we employ HyperOmni Vision for the present application.

In this research, omnidirectional images are acquired at video-rate by HyperOmni Vision placed in the sender side. First, HyperOmni Vision is fixed at a position with an appropriate position and height so that the



(a) Sensor (b) Layout design

Figure 1. HyperOmni Vision.

Figure 2. Omnidirectional image.

Figure 3. Common perspective image.

whole lobby or the whole conference room is contained in acquired omnidirectional images. Second, acquired omnidirectional images are converted to Digital Video stream. Finally, the converted stream is inputted to the sender computer.

2.3. Transportation of Omnidirectional Images

Our system uses Digital Video Transport System (DVTS)[14] for transportation of omnidirectional images through the internet. The DVTS is an application for transportation of IEEE1394 (FireWire) Digital Video stream through the internet and has the following five characteristics.

1. DVTS consumes about 33Mbps as network bandwidth to transport the high quality digital video stream (720 $\times$ 480 pixel, 30fps).
2. DVTS also transports auditory information synchronized with the visual scenes.
3. DVTS compresses images to IEEE1394 Digital Video stream and develops it with a DV recorder being the product for the consumer market.
4. DVTS can change the bandwidth adaptive for the end to end network condition.
5. Digital Video stream can be transported to anywhere through the internet.

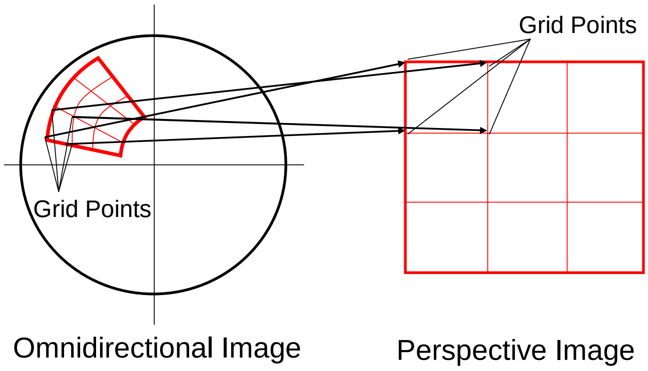


Figure 4. Relationship between omnidirectional image and perspective image.

2.4. Generation of Perspective Images

First, Digital Video stream sent by the sender computer is received by the receiver computer. Second this stream is developed to omnidirectional images with a DV recorder. Finally, view-dependent perspective images are generated by clipping a part of circular omnidirectional image.

In this research, image transformation is carried out in real-time, because it is very important that user's view-dependent perspective image is displayed to the user without the significant time delay. We actually generate uniformly spaced grids on a perspective image, and then compute the exact transformation only for sparsely spaced grid points for maintaining real-time computation.[6] The flow of view-dependent image generation from omnidirectional images is given in the followings.

1. At the beginning, user's head position and viewing direction are measured by using various interactive devices such as a mouse, joystick and 3D positioning sensors.
2. According to the above measurement, the corresponding viewing area in an omnidirectional image is determined and then clipped (see Figure 4). Note that the user's head position (viewpoint) is assumed to be located at the inner focal point O_M of the HyperOmni Vision.
3. A perspective image is computed from the clipped omnidirectional image and grid points by applying an image warping technique with bi-linear interpolation.[15] Note that the computed image is not exactly a perspective image but a pseudo-perspective image, and thus is geometrically distorted in some degree.

The above steps are iterated according to continuous measurements of user's view direction for providing an image sequence of virtual tour of a dynamic real scene. Moreover, in the multiple user mode, the above step is carried out for each user in the same way.

3. EXPERIMENT

We have implemented the internet telepresence system described in the previous section and have carried out two kinds of experiments: One is the intramural experiment through the campus LAN with 100Mbps Ethernet, and the other is the extramural experiment through the campus LAN and optical wireless LAN with 100Mbps.

3.1. Intramural Experiment and System Configuration

The system configuration and its hardware specification in the intramural experiment are illustrated in Figure 5 and Table 1. In the sender side, the captured omnidirectional images have the size of 720×480 pixels. In the receiver side, the received omnidirectional images and planar perspective images displayed on the face mounted display have the size of 720×480 pixels and 720×486 pixels (NTSC signals), respectively.

In the experiment, a sender computer is placed in the lobby [the grand floor] of Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), and a receiver computer is placed in our laboratory [the second floor]. Multiple users can look around the scene in different directions at the same time (see Figure 6).

Figure 7 shows two sequences among four users's sampled sequences of virtually looking around a real indoor scene which contains two walking persons in the lobby, as well as the corresponding sequence of input omnidirectional images. In this experiment, it can be seen that two persons appear in the scene in the period and views of users are individually tracking one of two walkers at the same time. The system provides multiple users with the natural feeling of looking around a real dynamic 3D environment at a fixed view point with auditory information. The time delay from the acquisition of omnidirectional image to the view-dependent perspective image display onto the face mounted display is about 1 sec. Digital Video stream (a sequence of omnidirectional images) is received every 33.3 msec without packet loss, and then the image display is also refreshed every 33.3 msec. The time delay from the change of user's viewing direction to the change of displayed image is 45 msec. It should be noted that the former time delay depends on the actual distance from the user (viewer) to the world to be seen and that the latter time delay does not depend on the distance, while conventional telepresence systems based on active regular cameras suffer from the significant time delay caused by the communication between both two sites.

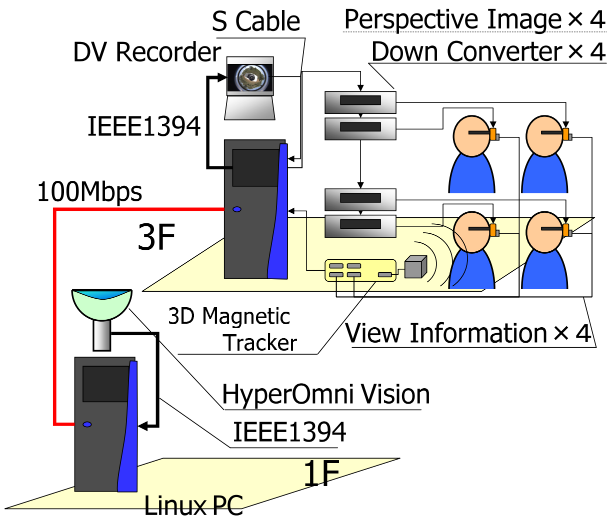


Figure 5. Hardware configuration of internet telepresence system in the intramural experiment.

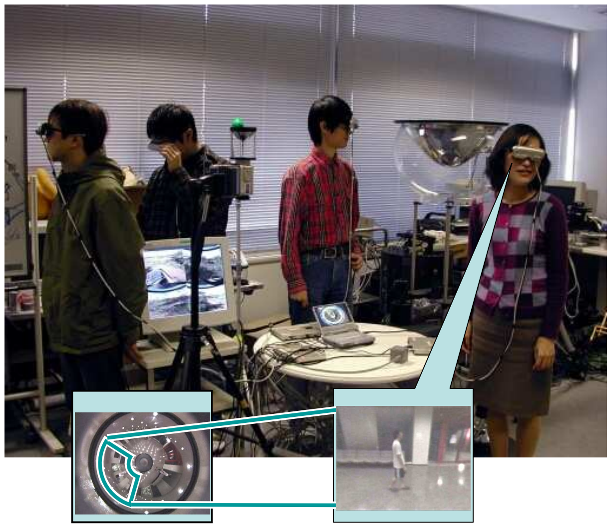
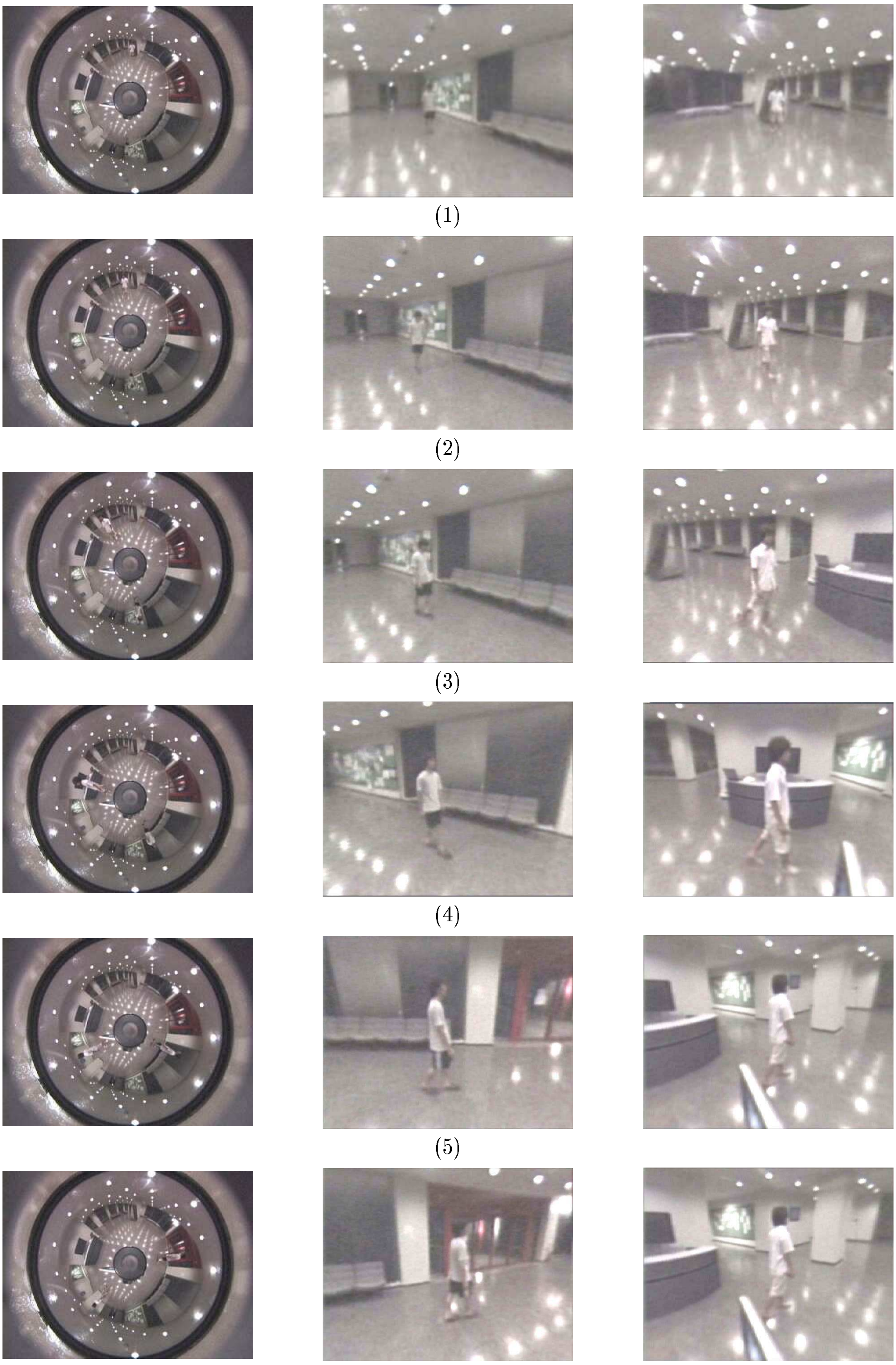


Figure 6. Appearance of four users in the intramural experiment.

Table 1. Specifications of internet telepresence system in the intramural experiment.

Sender and Receiver Computers	OS: Linux CPU: Pentium4 2GHz Memory: 512MB
Image Transport System	Digital Video Transport System $dvts - 0.9a22$
Image Display Device	OLYMPUS EyeTrek FMD-700
3D Magnetic Tracker	POLHEMUS 3SPACE FASTRAK



(a) Omnidirectional images in time sequence

(b) Generated looking-around sequence of images for user 1

(c) Generated looking-around sequence of images for user 2

Figure 7. Intramural experimental results.

3.2. Extramural Experiment and System Configuration

The system configuration and its hardware specification in the extramural experiment are illustrated in Figure 8 and Table 2. In this system, an immersive projection display with a four-facet multi-screen is employed as an image display device. Four planer perspective images displayed on the four-facet display have the size of 680 × 480 pixels.

In this experiment, a sender computer is placed in our laboratory, and a receiver computer is placed in TAO-Nara Research Center (TAO-NRC) located in the outside of NAIST. A user can look around the remote scene by controlling a joystick (see Figure 9).

Figure 10 shows experimental results in the long distance using the network which contains the optical wireless LAN. In this experiment, the time delay from the acquirement of omnidirectional image to the perspective image display onto the screen is about 1 sec. Digital Video stream is received every 33.3 msec without packet loss, and then the image display is also refreshed every 33.3 msec. While perspective images controlled by user's joystick are generated in real-time with very little time delay, the time delay from the change of user's viewing direction to the change of displayed image does not arise. Thus this could promises that we realize the real-time telepresence in the long distance through the internet. Moreover, by projecting perspective images to a variety of displayed device, this could lead to promising applications such as virtual sightseeing, teleconferencing and lesson, and so on.

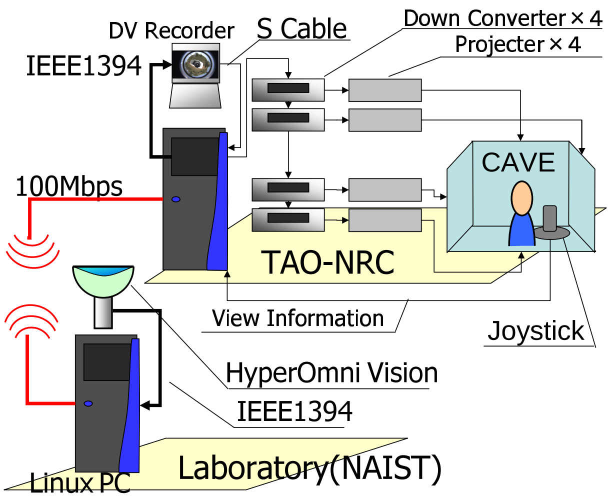


Figure 8. Hardware configuration of internet telepresence system in the extramural experiment.

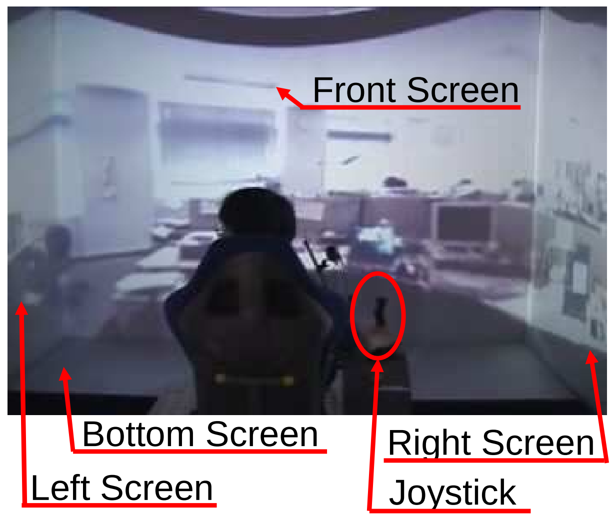
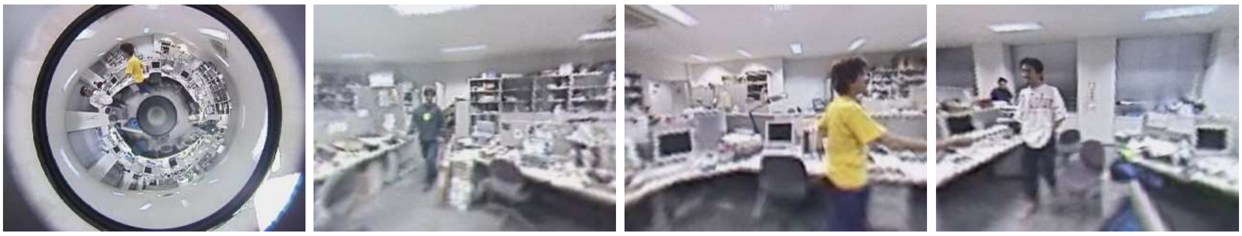


Figure 9. Appearance of a user in the extramural experiment.

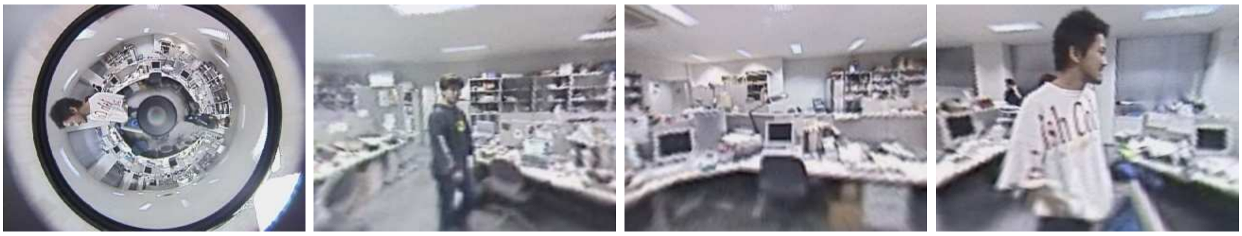
Table 2. Specifications of internet telepresence system in the extramural experiment.

Sender and Receiver Computers	OS: Linux CPU: Pentium4 2GHz Memory: 512MB
Image Transport System	Digital Video Transport System $dvts - 0.9a22$
Image Display Device	Four Faced Display (CAVE)

(1)

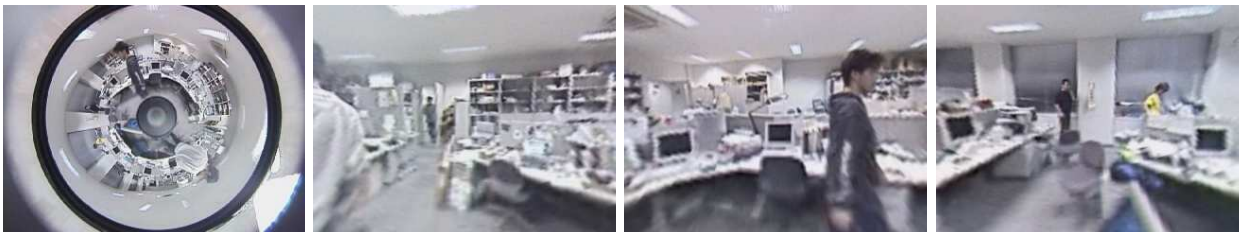


(2)

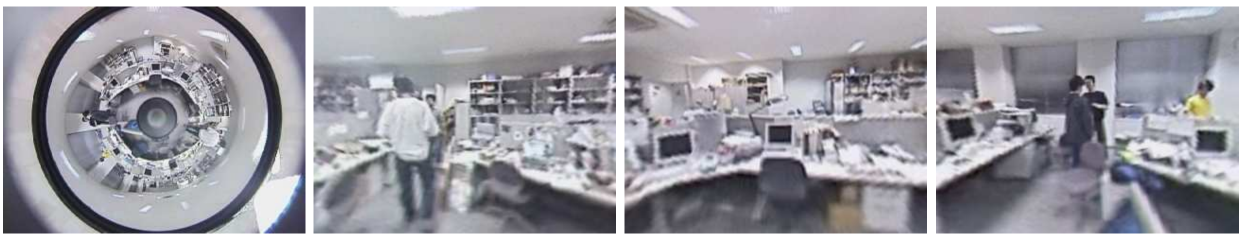


(3)

(4)



(5)



(6)

(a) Omnidirectional images in time sequence
(b) Generated sequence of images for left screen
(c) Generated sequence of images for front screen
(d) Generated sequence of images for right screen

Figure 10. Extramural experimental results.

4. DISCUSSION

Experimental results have exhibited the satisfactory performance of the developed telepresence system. The real-time internet telepresence system enables us to tour into a visualized dynamic real world that is far from the user's site in real-time through the internet. In other words, we can virtually observe a "mirrored" world which varies in concurrence with corresponding real 3D environment through the internet.

It has been experimentally proven that the proposed telepresence system provides us with the natural feeling of looking around a real dynamic world. However, there still remain some problems; the first problem is that the quality of images displayed on a face mounted display is low, especially in an area of scene which is projected around the center of the hyperboloidal mirror of omnidirectional video camera. This comes from the intrinsic property of HyperOmni Vision. This could be resolved by using a high-resolution omnidirectional video camera based on different optics. The second problem is the blurring in some parts of the image, which is caused by the varying depth of field on the curved mirror surface of HyperOmni Vision. This could be relaxed to some extent by adjusting the lens aperture. These should further be investigated. The third problem is that the network bandwidth consumed to transport omnidirectional images through the internet is largish. This makes it impossible to transport omnidirectional images through the general network such as ADSL, 10Mbps Ethernet, and 11Mbps wireless LAN. The fourth problem is that omnidirectional images can not be transported to multiple sites. The latter two problems could be resolved by replacing DVTS with Windows Media.

5. CONCLUSIONS

This paper has described a new system of networked telepresence which realizes virtual tours into a visualized dynamic real world on the internet. The proposed telepresence system is characterized by real-time omnidirectional image acquisition, omnidirectional image transportation through the internet and real-time generation of user's view-dependent images from dynamic omnidirectional images. Experiments have shown that the time delay from the change of viewing direction to the change of displayed image is small and does not depend on the actual distance between the sender and receiver sites and multiple users can look around from a single viewpoint a visualized dynamic real world that is far from an observation site in different directions at the same time through the internet. As a practical application, we expect to apply this telepresence system to tele-operation, teleconferencing and lesson, virtual tour, remote surveillance, and aquarium telepresence.

The future work includes: (1) application based on using a high-resolution omnidirectional video camera; (2) relaxation of the blurring in some parts of the image; (3) for more practical applications, reduction of network bandwidth consumed in order to transport Digital Video stream; (4) extension toward a multicast system in which omnidirectional images are simultaneously transported to multiple sites; (5) development of bidirectional communication using auditory information between the sender and receiver sides. These could drastically improve the reality of telepresence in the future.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Hideki Sunahara, Nara Institute of Science and Technology, for his helpful comments. We also thank Mr. Akimichi Ogawa, University of Keio, and Miss. Michiko Izumi, Nara Institute of Science and Technology, for their cooperation and helpful comments during this work.

REFERENCES

1. N.Greene: "Creating raster omnimax images from multiple perspective views using the elliptical weighted average filter", IEEE Computer Graphics and Applications, Vol.16, No.6, pp.21-27, 1986.
2. A.Krishnan and N.Ahuja: "Panoramic image acquisition", Proc. 1996 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.379-384, 1996.
3. R.Szeliski: "Video mosaic for virtual environments", IEEE Computer Graphics and Applications, Vol.16, No.2, pp.22-30, 1996.
4. J.Y.Zhebg and S.Tsuji: "Panoramic representation of scenes for route understanding", Proc. 10th Int. Conf. on Pattern Recognition, Vol.I, pp.161-167, 1990.

5. V.Peri and S.K.Nayar: "Generation of perspective and panoramic video from omnidirectional video", Proc. DARPA Image Understanding Workshop, pp.243-245, 1997.
6. 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, Vol.71, No.2, pp.154-165, 1998.
7. K.Der, A.Basu and J.Reyda: "Interactive viewing of panoramic images", Proc. Vision Interface '97, pp.162-169, 1997.
8. S.K.Nayar and T.E.Boult: "Omnidirectional VSAM systems: PI-report", Proc. DARPA Image Understanding Workshop, pp.55-61, 1997.
9. Z.L.Cao, S.J.Oh and E.L.Hall: "Dynamic omnidirectional vision for mobile robots", Journal of Robotic Systems, Vol.3, No.1, pp.5-17, 1986.
10. Y.Yagi, S.Kawato and S.Tsuji: "Real-time omnidirectional image sensor (COPIS) for vision-guided navigation", IEEE Trans. on Robotics and Automation, Vol.10, No.1, pp.11-22, 1994.
11. J.Hong, X.Tan, B.Pinette, R.Weiss and E.M.Riseman: "Image-based homing", Proc. 1991 IEEE Int. Conf. on Robotics and Automation, pp.620-625, 1991.
12. K.Yamazawa, Y.Yagi, and M.Yachida: "New real-time omnidirectional image sensor with hyperboloidal mirror", Proc. 8th Scandinavian Conf. on Image Analysis, Vol.2, pp.1381-1387, 1993.
13. K.Yamazawa, Y.Yagi, and M. Yachida: "Obstacle detection with omnidirectional image sensor HyperOmni Vision", Proc. 1995 IEEE Int. Conf. on Robotics and Automation, pp.1062-1067, 1995.
14. A.Ogawa, K.Kobayashi, O.Nakamura, and J.Murai: "Design and Implementation of DV Stream Over Internet", IWS Internet Workshop, 1999.
15. G.Wolberg: Digital Image Warping, IEEE Computer Society Press, 1990.