Server and Client Simulator for Web-based 3D Image Communication

Jung-Hwan Ko, Sang-Tae Lee, and Eun-Soo Kim*

Abstract

In this paper, a server and client simulator for the web-based multi-view 3D image communication system is implemented by using the IEEE 1394 digital cameras, Intel Xeon server computer and Microsoft's DirectShow programming library. In the proposed system, two-view image is initially captured by using the IEEE 1394 stereo camera and then, this data is compressed through extraction of its disparity information in the Intel Xeon server computer and transmitted to the client system, in which multi-view images are generated through the intermediate views reconstruction method and finally display on the 3D display monitor. Through some experiments it is found that the proposed system can display 8-view image having a grey level of 8 bits with a frame rate of 15 fps.

Keywords : 3D communication, 3D imaging and display, multi-view image

1. Introduction

Recently, many research works have been actively done on 3D imaging and display technology due to its great demand throughout the world. The real 3D display system of the future should have a full-color, full-parallax 3D imaging capability. These advances will make viewers feel as if they are actually in the real scene itself so-called the tele-presence [1].

Depending on the development stages of the 3D imaging & display technologies, various degrees of resolution and spatial perception are now offered. However, most of the developed 3D display systems, require the viewer to wear special glasses, which is the main shortcoming deterring them from gaining wide acceptance. Therefore, many researchers have been conducting in-depth studies on the auto-stereoscopic 3D display system as it can present 3D image to its viewers without the need to wear any special glasses.

In the 3D image communication system, as the number of views increase, the system complexity in hardware and software also tend to simultaneously increase, making practical implementation of the multi-view 3D image communication system very difficult. In addition, most of the conventional multi-view stereoscopic 3D display system has been developed for 3DTV or 3D monitor applications, but only a few research works have been done for the web-based multi-view 3D imaging and display applications [2-4].

In general, humans can recognize their different viewpoint images of an object through their eyes and the human brain recognizes 3D stereopsis of an object by synthesizing it with its difference, i.e., the binocular disparity, of stereo input image. The conventional 3D imaging and display system has been implemented by imitating the principle of the human visual system (HVS).

A web-based 3D imaging and display system is largely composed of two parts; server and client system. In the server system, a two-view image of the object is captured by a pair of web camera and compressed by the disparity data extracted from these images and transmitted to the client. In the client system, multi-views for the object are generated by using the intermediate views reconstruction (IVR) method in order to display the 3D image more naturally by using the received two-view image and loaded on the 3D display system so that the viewers can experience the displayed image three dimensionally.

In this paper, a server and client simulator for the web-

Manuscript received October 22, 2004; accepted for publication December 15, 2004.
This works have been conducted by the ITRC project of the Ministry of Information and Communication of Korea and the NRL project of the Ministry of Science and Technology of Korea.
* Members, KIDS
Corresponding Author : Jung-hwan Ko
Dept. of Electronic Eng., Kwangwoon University 447-1 Wolge-Dong, Nowon-Gu, Seoul 139-701, Korea.
E-mail : misagi@explore.kw.ac.kr **Tel** : +2 940-5118 **Fax** : +2 941-5979

based multi-view 3D image communication system for teleconferencing and telemedicine applications is designed and implemented by using the IEEE 1394 digital cameras, Intel Xeon server computer and Microsoft's DirectShow programming library and its performance was analysed in terms of image-grabbing frame rate, displayed image resolution and number of views. A feasibility test for the proposed system was also carried out by implementing the eight-view 3D image communication system with a grey level of 8 bits and a frame rate of 15 fps.

2. Web-based 3D Image Communication System

As shown in Fig. 1 the proposed web-based 3D communication system is composed of two parts, namely the server and client parts. In the server part, two IEEE 1394 digital cameras are used for capturing two-view images of an object. These cameras can be controlled through the on-line computer system, such that no other additional capturing device that is normally required in most conventional camera systems, are needed. The

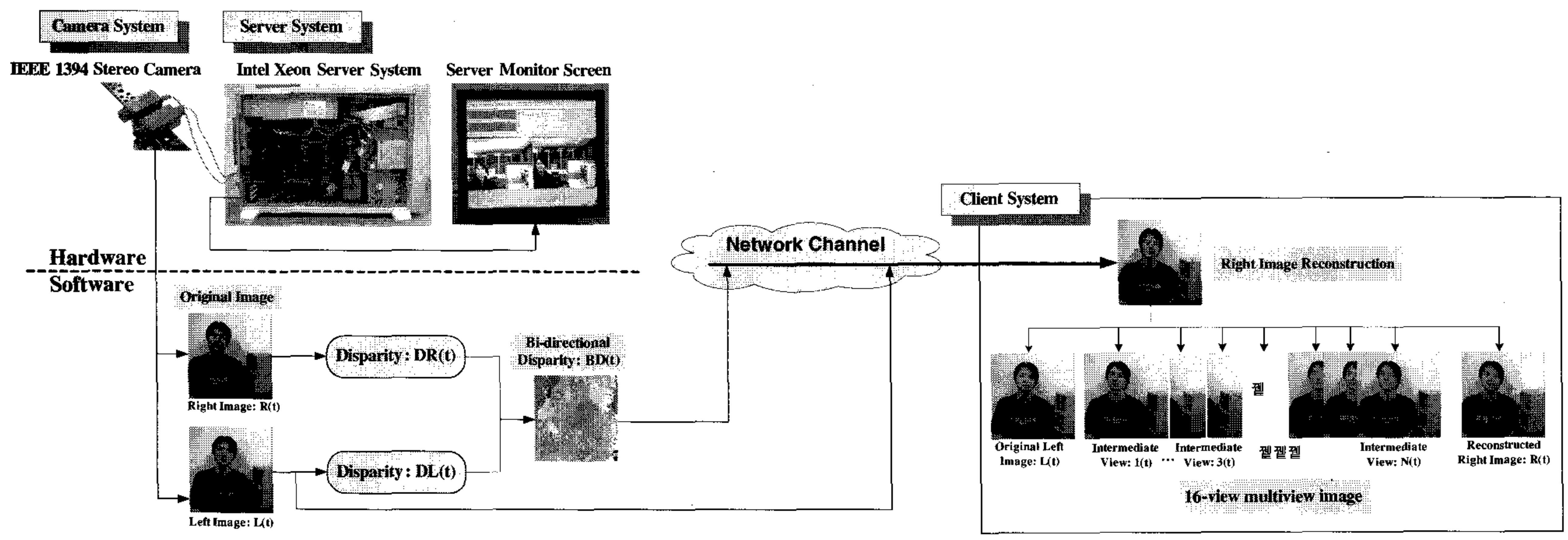

Fig. 1. Overall web-based 3D communication system.

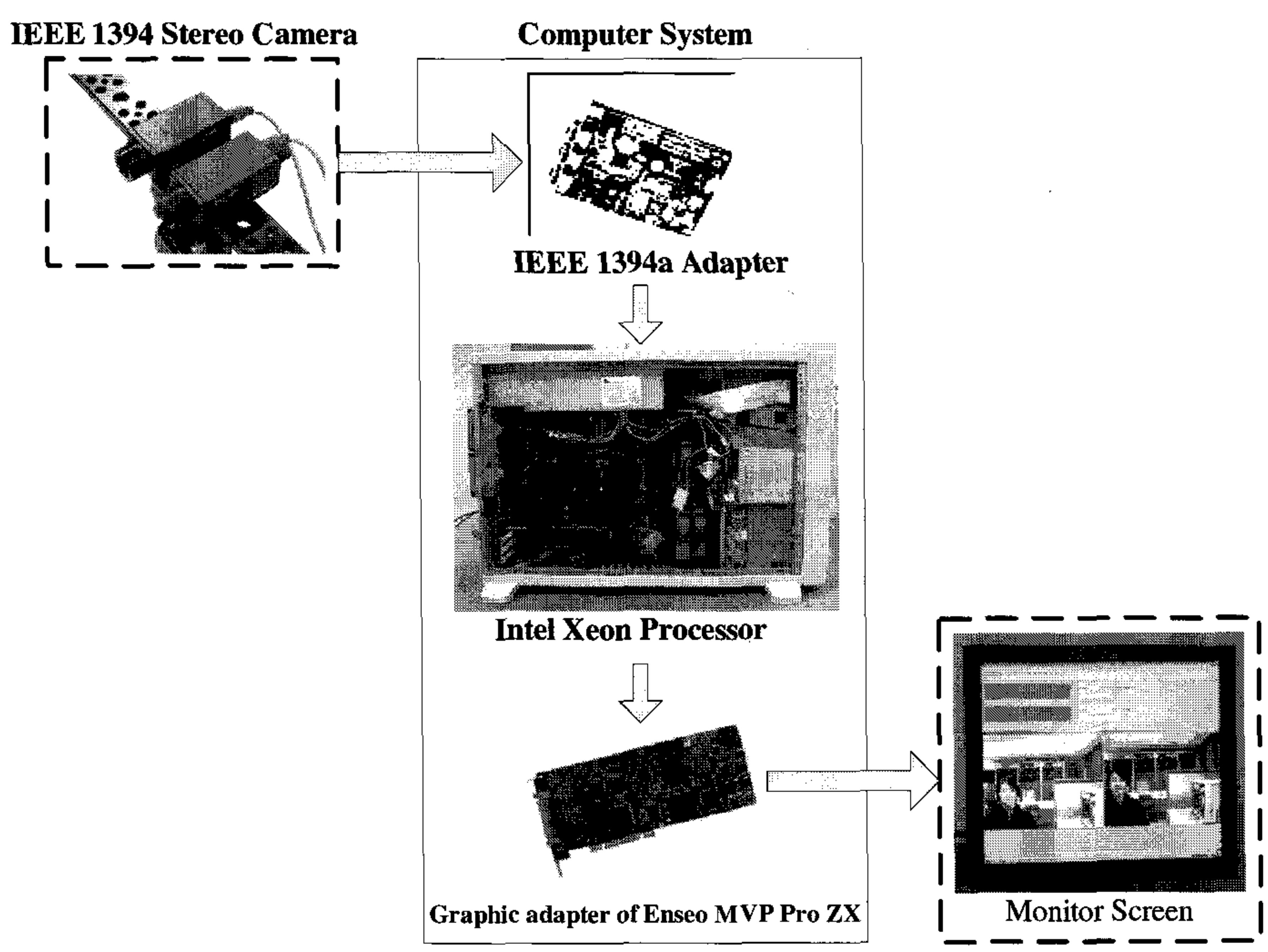

Fig. 2. Block-diagram of the proposed system.

proposed 3D communication system operates as follows: The captured two-view image is processed in the Intel Xeon server computer system, and then, the processed data is grabbed by a graphic card having two ports for supporting two-view stereoscopic display system in real time. These outputs are transmitted to the client system through the communication network, in which the received two-view data is reconstructed, and based on this data an eight-view image is generated through the IVS method [5] to display a 3D image, in a more natural manner through client system and then, finally loaded on to the practical 3D display system.

In this paper, a program to control the overall system is developed using the Microsoft's DirectShow programming library, and two Unibrain Fire-i400 digital cameras having a built-in A/D converter are used as a powerful and flexible stereo camera system. Since in this camera system the IEEE 1394a is employed as a camera interface, this system can support for data trans-mission of 400 Mbps (=50 MB/s). Moreover, the camera parameters such as gain, shutter, black level, gamma, backlight compensation, sharpness, white balance and color saturation can be easily controlled on the software basis such that in the proposed camera system, the operational difference between the stereo camera can be resolved by adjusting their parameters through the digital programs. Thus, with this digital camera system, we can obtain better quality multi-view images in the same environments.

Moreover, the computer system is implemented based on the Intel Xeon, which is largely composed of a dual Intel Xeon processor of 2.4 GHz, dual 512 M DDR Memory, server board SE7501HGZ and SCSI ultra320 HDD as shown in Fig. 2. As can be seen in Fig. 2, the camera system can transmit data at the rate of 200 MB/s at maximum, but in the case of IEEE 1394 adapter, it can only receive them at the rate of 132 MB/s at maximum, because it must transmit data while taking into account the maximum bandwidth of the computer input system. The computer system has PCI-X buses (two slots with the rate of 800 MB/s, one slot with the rate of 1056 MB/s), but the IEEE 1394 adapter is only compatible with the PCI bus (one slot with the rate of 132 MB/s) not the PCI-X bus so that, the bus can transmit data at a maximum rate of 132 MB/s. Similarly because the graphic adapter also can't support the PCI-X bus, it might transmit data at a rate of 132 MB/sec in maximum.

The server board, SE7501HGZ, has 3 separate PCI buses, namely three 32 bit/ 33 MHz PCI, two 64 bit/ 100 MHz PCI-X and one 64 bit/ 133 MHz PCI-X. The 32 bit/ 33 MHz PCI is a general PCI bus and it can support 132 MB/sec. The 64 bit/ 100 MHz PCI-X and 64 bit/ 133 MHz PCI-X can support 800 MB/s and 1056 MB/sec, respectively. All of these buses work separately such that data flow of one bus does not interfere with those of other buses. These bus characteristics make it possible to transmit data from the camera to the computer system at a high data rate.

In general, in order to transmit data at high speed or to operate many image processing algorithms, the computer system should have a powerful CPU and enough memory to process them. Inherently, the Intel Xeon server system is designed for the server system and not for the general-purpose desktop computer system. Thus, it is very suitable for high data transmission, heavy process of algorithms and it has a reliable stability as well. Moreover, in this computer system, the SCSI ultra320 HDD is used as a data storage system, and an operating system of Microsoft Windows 2000 server is installed in the main data storage system. By using the IVS technique [5], more natural stereo stereopsis can be acquired by synthesizing the multi-view stereo images from the limited given stereo image. As a result, stereopsis can be offered to many observers by implementing of available multi-view 3D display system.

Fig. 3 shows a general concept of intermediate views synthesis, in which 'L', 'R' and 'I' represent the left and right and intermediate image planes. The position of the corresponding point of the intermediate view image is defined as a normalized distance from the left image. The distance from the left to the right image plane is normalized to be '1', such that $\alpha \in [0,1]$. For example, if $\alpha = 0$ is the left image and $\alpha = 1$ is the right image and the interval values mean the intermediate view image.

As shown in Fig. 3, based on relation such as Eq. (1), point $I_I(i, j)$ of the intermediate view image plane can be reconstructed from disparity vector $\hat{d}_{ij}$ of the left, right and distance α as shown in Eq (2).

$$\hat{d}_{ij} = d_L + d_R = \alpha \hat{d}_{ij} + (1-\alpha)\hat{d}_{ij} \tag{1}$$

$$I_I(i, j) = I_L(i + \alpha \hat{d}_{ij}, j) = I_R(i - (1-\alpha)\hat{d}_{ij}, j) \tag{2}$$

Where $\hat{d}_{ij}$ is the disparity value in search range, I_p is the intersection that disparity of the left image and right image is crossed, I_L is the block of the left side image and I_R is the block of the right side image corresponding to I_L. In this paper, more natural intermediate views can be reconstructed by using interpolation scheme with a weighted average as shown in Eq. (3) which represents a case of interpolation with a weighted average by position α of viewpoint.

$$I_P(i,j)=(1-\alpha)\cdot I_r(i-\hat{d}_{ij}(i,j),j)+\alpha\cdot I_l(i-\hat{d}_{ij}(i,j),j) \quad (3)$$

Generally, during the intermediate view reconstruction process some occluded regions may form where one of the stereo cameras sees while the other does not and some regions the allocation of the disparity vector overlap. Therefore, in this paper, the disparity vectors for these regions are substituted with the mean values of the disparity vectors of the nearby regions by disparity regularization [6, 7]. If the viewpoint is not occluded, then the disparity is defined as the distance between the image points in both images. Eq. (4) shows the disparity in the horizontal direction and the relationship between left and right images, where I_R, I_L represent the right and left images at coordinates $(i_R + j_R)$ and $(i_L + j_L)$, respectively.

$$I_R=\begin{bmatrix} i_R \\ j_R \end{bmatrix}=\begin{bmatrix} i_L+\hat{d}\ (i_L,j_L) \\ j_L \end{bmatrix}=I_L+\begin{bmatrix} \hat{d}(i_L,j_L) \\ 0 \end{bmatrix} \quad (4)$$

Accordingly, through the intermediate views synthesis, multi-view image can be created and displayed three-dimensionally on the practical 3D display system in a more natural manner.

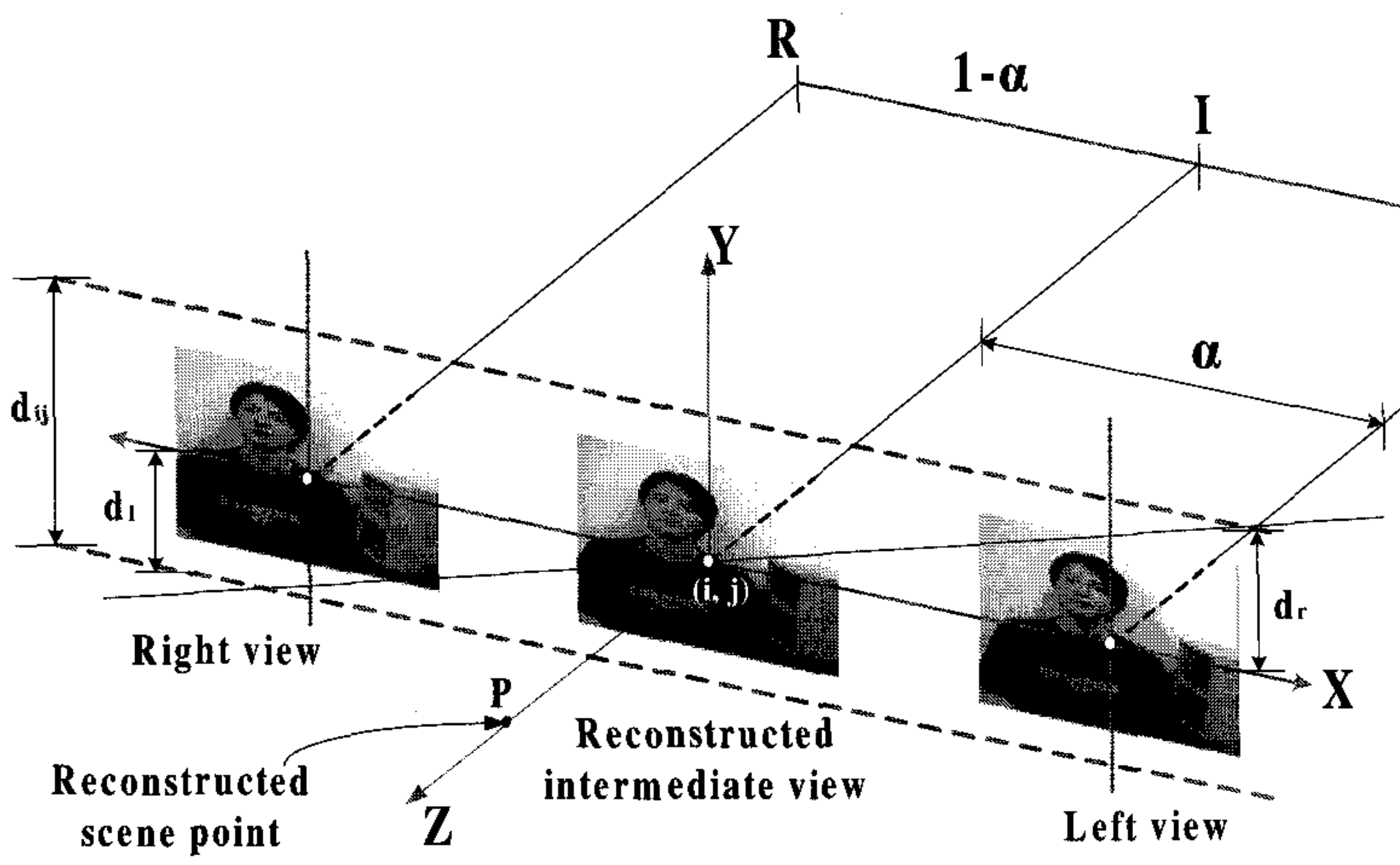

Fig. 3. Corresponding point of intermediate view from left and right image.

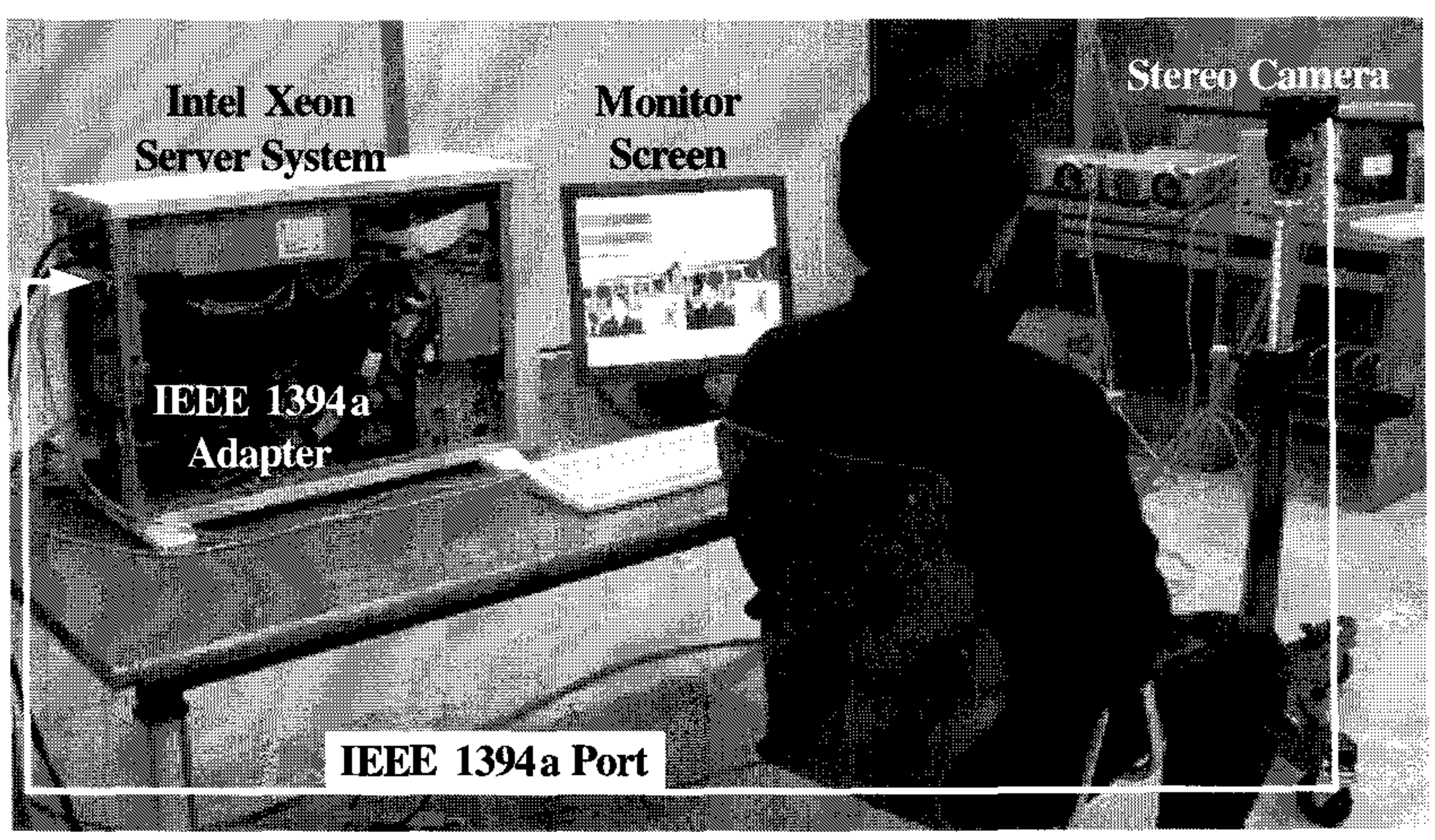

Fig. 4. Experiment setup for the proposed 3D image communication.

3. Experiments and its Results

Fig. 4 shows the experimental set-up for the proposed real-time web-based 3D image communication system that consists of IEEE 1394 digital camera system, Intel Xeon server system and display system. Fig. 5 shows the server simulator to evaluate the performance of the proposed system. Each digital camera system captures image data at the rate of 15 frames/sec with a resolution of 640 × 480 pixels. In the computer simulation, two-view images of 640 × 480 pixel were captured from the stereo camera system, as shown in Figs. 6 (a) and (b), and Fig. 6 (c) shows the disparity maps extracted from these images by using the adaptive disparity estimation algorithm [5]. Here, the forward and reverse disparity maps were generated from the captured images and bi-directional disparity map was produced from two disparity maps. Then the original left image and the bi-directional disparity map were transmitted to the client through the communication channel. In the experiment, the stereo image pairs of 30 frames captured from the digital cameras for 2 seconds were used for generating eight-view images using the IVR scheme at the client side.

Fig. 7 shows the implemented client simulator, in which the received left image and disparity data and the right image reconstructed from them are displayed on the top left of the simulator screen.

Table 1 shows the synthesis processing time of intermediate image as view of the proposed 3D imaging communication system.

Table 2 shows the peak-signal-to-noise (PSNR) of the intermediate view synthesized by each algorithm for the images. PSNR is computed by using Eq. (5), in which the root mean squared error (RMSE) means the square root of MSE.

$$PSNR = 20\log_{10}\left[\frac{255}{RMSE}\right] \tag{5}$$

The number of views, frames per second and delay time per frame are displayed at the top right of the simulator screen. In addition, the generated intermediate sixteen-view images are displayed at the lower part of the simulator screen. Since the IEEE 1394 adapter can transmit 50 MB/sec per port and up to 200 MB/sec per 4 ports, there is no problem in transmission of the camera data. The computer system has three kinds of PCI buses of which the bandwidths are 132 MB/sec, 800 MB/sec, and 1056 MB/sec, respectively. However the IEEE 1394a adapter

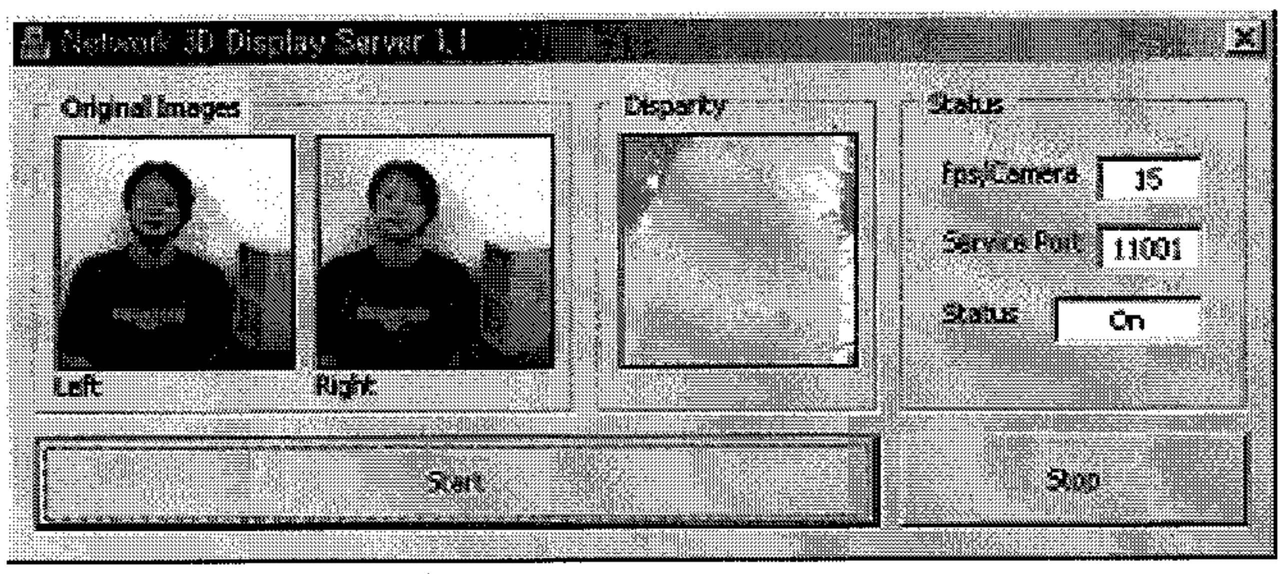

Fig. 5. Server simulator system.

(a) Left image

(b) Right image

(b) Disparity map

Fig. 6. Captured 2-view images and its disparity map.

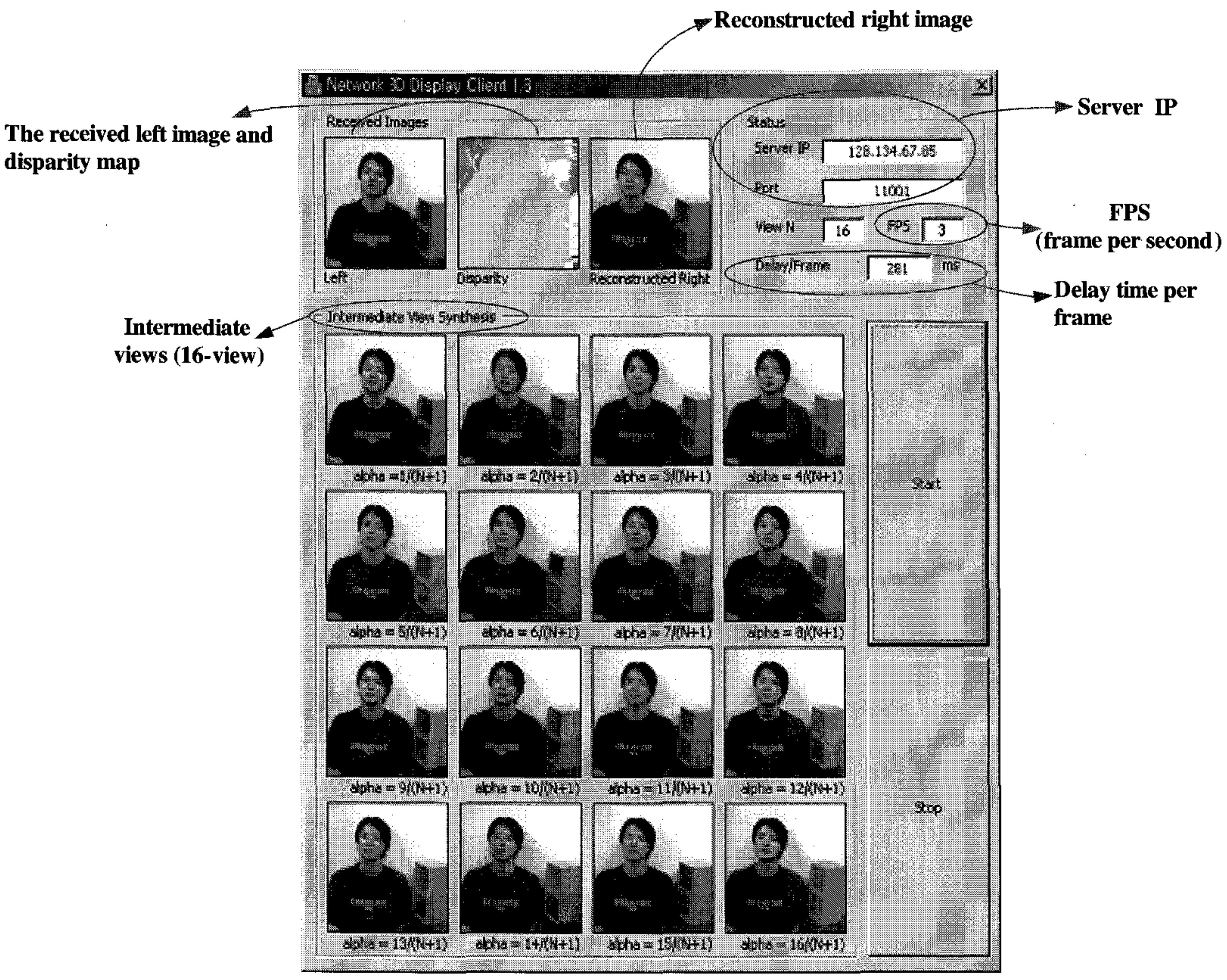

Fig. 7. Client simulator for 16-view 3D image communication

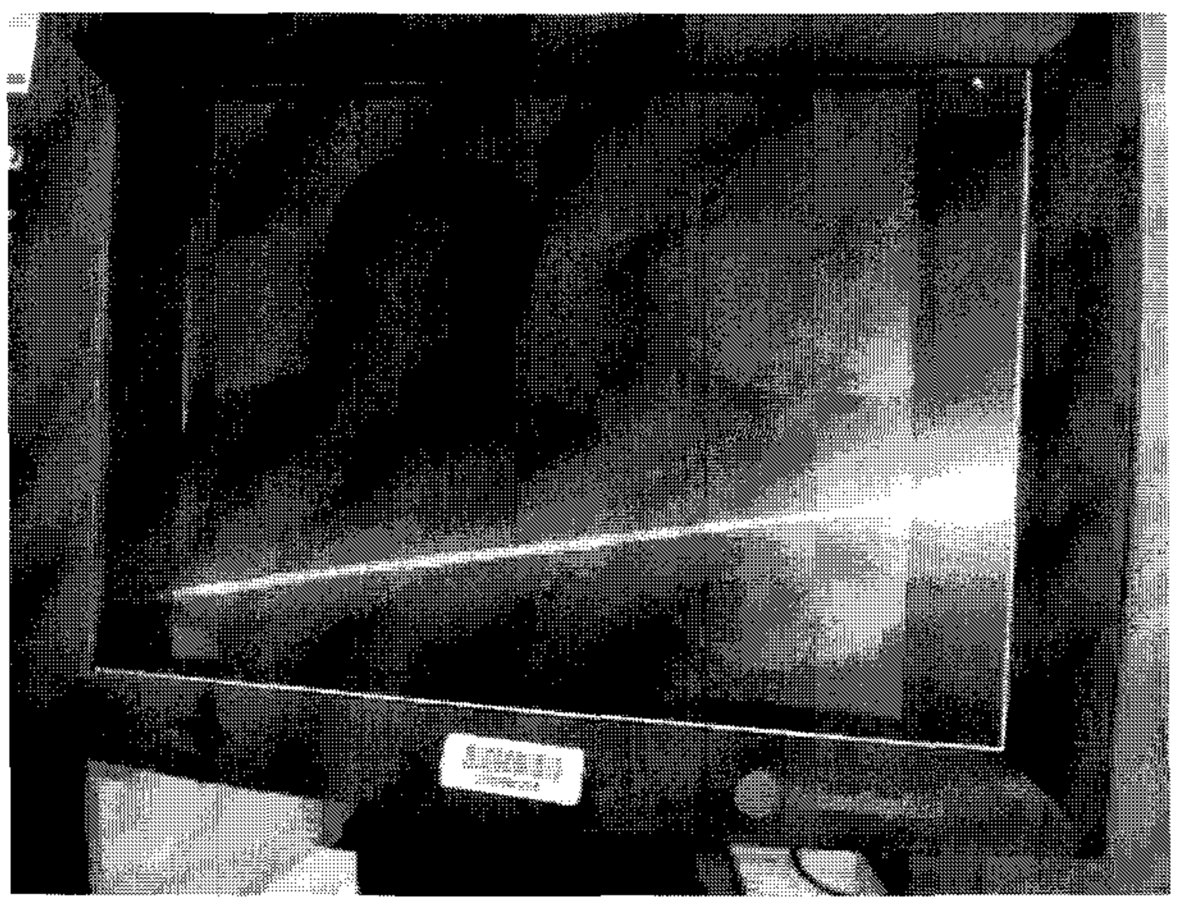

Fig. 8. 8-view image for the test image of 'Man' on the commercial 3D display monitor of *SiliconGraphics*.

and the graphic adapter can support only the PCI bus, and not the PCI-X bus, such that the data rate of the PCI bus is limited to 132 MB/sec. Therefore, in the proposed system, all the PCI buses work as the general PCI devices.

The graphic adapter of *Enseo MVP Pro ZX* has a PCI bus and four video-output ports as interface, so that it can support four display outputs. The graphic adapter used in this paper can support 8 MB memory per port and 32 MB memory per adapter. From the experimental results, it is found that the display adapter can support an eight-view display at the rate of 15 frames/sec having a resolution of 256×256 pixels with a grey level of 8 bits. But the frame

Table 1. Synthesis processing time of intermediate image as view

Experiment	Process time [unit: ms]		
	1 view	4 view	16 view
1	30	70	281
2	31	67	276
3	30	63	280
4	30	66	280
5	30	70	277
Total average	30.2	67.2	278.8

Table 2. PSNR results between original right image and reconstructed right image

Frame	Simulation result
1st	30.470494 dB
5th	30.347301 dB
10th	30.251242 dB
Total average	30.356345 dB

rate and number of views are mutually interrelated and can be easily controlled in the proposed system by using the developed software program. Accordingly, greater flexibility in design of the proposed multi-view 3D display system can be expected in the practical applications. Figure 8 show the 8-view image for the test image of 'Man' on the commercial 3D display monitor of *SiliconGraphics 1600 SW*. This monitor uses multiple perspective views and a *microlens* array (or *lenticular* screen) in juxtaposition with a flat panel display screen to achieve a stereoscopic depth effect. These experimental results clearly shows that it is possible to implement a new real-time web-based remote multi-view 3D imaging communication system using the proposed scheme finally suggested.

4. Conclusion

In this study, a server and client simulator for the web-based multi-view 3D image communication system was implemented by using the IEEE 1394 digital cameras, Intel Xeon server computer and Microsoft's DirectShow programming library. Through the experiments conducted in this research, it was found that the proposed system can display 8-view image having a grey level of 8 bits with a frame rate of 15 fps. The feasibility test for the proposed system also showed that it is possible to implementate this web-based real-time multi-view 3D image communication system.

References

[1] N J. R. Moore, N. A. Dodgson, A. R. L. Travis, and S. R. Lang, in *Proc. of SPIE* (1996), **2653**, p. 10.

[2] M. Ebroul Izquierdo, *Signal Processing: Image Communications* **11**, 231 (1998).

[3] M. Siegel, in *Proc. of SPIE* (1994), **2177**, p. 258.

[4] V. Grinberg, in *Proc. of SPIE* (1994), **2177**, p. 56.

[5] K. H. Bae, J. J Kim, and E. S. Kim, *Optical Engineering*, **42**, 1778 (2003).

[6] F. Heitz and P. Bouthemy, *IEEE Trans. PAMI*, **15**, 1217 (1993).

[7] H.H. Nagel and W. Enkelmann, *IEEE Trans, PAMI*, **8**, 565 (1986).

[8] http://www.sgi.com.