Web-based Real-time 3D Video Communication System for Reality Teleconferencing

Jung-Hwan Ko, Dong-Kyu Kim, Dong-Chun Hwang, Eun-Soo Kim

3D Display Research Center/ National Research Lab. of 3D Media, School of Electronic Eng., Kwangwoon University, 447-1Wolge-Dong, Nowon-Gu, Seoul 139-701, Korea

Abstract

In this paper, a new multi-view 3D video communication system for real-time Reality teleconferencing application is proposed by usin gthe IEEE 1394 digital cameras, Intel Xeon server computer system and Microsoft's DirectShow programming library and its performance is analyzed in terms of image-grabbing frame rate and number of views. The captured two-view image data is compressed by extraction of disparity data between them and transmitted to another client system through the communication network, in which multi-view could be synthesized with this received 2-view data using the intermediate view reconstruction technique and displayed on the multi-view 3D display system. From some experimental results, it is found that the proposed system can display 16-view 3D images with a gray of 8bits and a frame rate of 15fps.

1. Introduction

Recently, many research works are actively being done on 3D video communication technology due to its high interests throughout the world [1]. In the conventional multi-view 3D imaging communication system, as the number of views increases, the system complexity in hardware and software also tends to increase at the same time, so that it makes the practical implementation of the multi-view 3D imaging communication system so difficult until now. In addition, most conventional multi-view display systems have been under developing for 3DTV or 3D monitor applications, but there are a few research works done for the web-based 3D teleconferencing applications.

In general, multiple cameras used for capturing muitiview images of an object are required to have the mutually same operational characteristics. But, some operational differences between them might happen to be occurred in the conventional camera system so that, a process of adjusting those camera parameters are needed. Because in the most conventional multiview 3D imaging and display systems, the analog camera system has been employed in which the camera parameters are manually controlled and as a result in the practical experiments, it is very difficult to keep all of the used cameras in same operational conditions. In order to resolve this problem, some additive hardware devices for adjusting the camera parameters and transmitting the captured images to them are required in the analog camera system.

Accordingly, in this paper, a new multi-view 3D video communication system for reality teleconferencing application is proposed by using the IEEE 1394 digital cameras, Intel Xeon server computer system, graphic card having multiple outputs and Microsoft DirectShow SDK. A feasibility test for the proposed system is also carried out by implementing the 16-view 3D video communication system with a gray of 8 bits and a frame rate of 15 fps.

2. Proposed Web-based Real-time 3D Video Communication System for Reality Teleconferencing

As shown in Fig.1 the proposed 16-view 3D communication system is composed of three parts such as the camera, server and client parts. In the camera part, two IEEE 1394 digital cameras are used for capturing two-view images of an object.

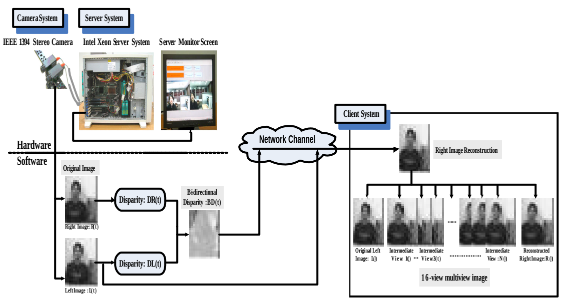

Fig. 1 Block-diagram of 3 parts composing the proposed system; camera, server and client parts

In this paper, the computer system is implemented basing-on the Intel Xeon, which is mainly consisted of dual Intel Xeon processor of 2.4 GHz, dual 512M DDR Memory, server board SE7501HGZ, SCSI ultra320 HDD as shown in Fig. 2.

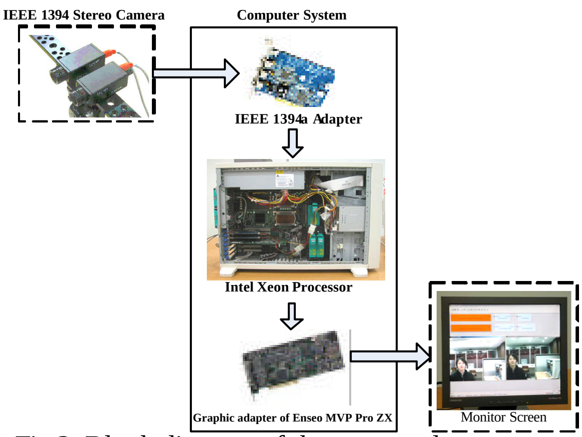

Fig 2. Block-diagram of the proposed system

In Fig. 2, the camera system can transmit data at the rate of 200MB/s in maximum. But the IEEE 1394 adapter can receive them only at the rate of 132MB/s in maximum because it must transmit data considering the maximum bandwidth of the computer input system. The computer system has the PCI-X buses (two slots with the rate of 800MB/s, one slot with the rate of 1056MB/s), but the IEEE 1394 adapter is only compatible with the PCI bus (one slot with the rate of 132MB/s) not the PCI-X bus so that, the bus can transmit data at a maximum rate of 132MB/s. Similarly because the graphic adapter also can't support the PCI-X bus, it might transmit data at a rate of 132MB/sec in maximum.

Also, by using the IVS technique [5], more natural stereo stereopsis can be acquired through synthesizing the multi-view stereo images from the limited given stereo image. As a result, offering of stereopsis to many observers is possible by implementation of available multi-view 3D display system [8]. Fig 4 shows a general concept of intermediate view synthesis. Fig. 3 is intersected the desired coordinate position in the intermediate image. It is corresponded from the left image plane 'L' to the right image plane 'I' through the plane of intermediate view image 'I' [9]. The position of corresponding point the intermediate view image is defined as a normalized distance a from the left image.

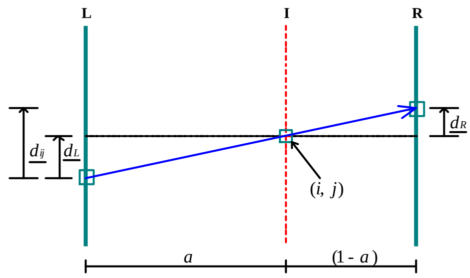

Fig. 3. Corresponding point of intermediate view image from left and right image

The distance from the left to the right image plane is 1, such that $a \in [0,1]$. For example, if $a = 0$ is the left image and $a = 1$ is the right image and the interval values mean the intermediate view image.

$$\hat{d}_{ij} = d_L + d_R = a\,\hat{d}_{ij} + (1-a)\,\hat{d}_{ij} \qquad (1)$$

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

$\hat{d}_{ij}$ is disparity value in search range, I_p is intersection that disparity of left image and right image is crossed, I_L is block of left side image, I_R is block of right side image corresponded to I_L. As shown in Fig. 4, because of relation such as Eq. (1), usually, point $I_I(i, j)$ of the intermediate view image plane can be reconstructed from disparity vector $\hat{d}_{ij}$ of left, right and distance a with Eq (2). Also, in this paper, more natural intermediate views can be reconstructed by using interpolation scheme with a weighted average. Eq. (3) shows the case of interpolation with a weighted average by position a of viewpoint.

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

Accordingly, through this process of intermediate views synthesis, multi-view 3D image can be created and displayed as more natural 3D image.

3. Experiments and Discussions

Figure 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.

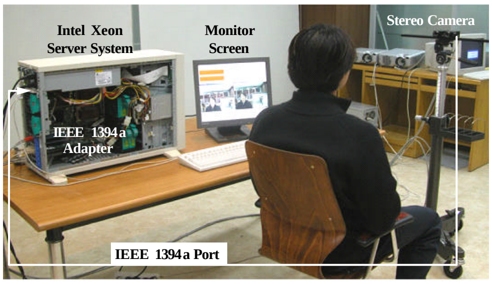

Fig. 4 Experiment setup for the proposed 3D image communication

Figure 5 shows the server simulator for performance evaluation of the proposed system.

Fig. 5 Server simulator system

That is, 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, 2-view images of 640×480 pixel are captured from the stereo camera system, which are shown in Fig. 6 (a), (b) and Fig. 6 (c) shows the disparity maps extracted from them by using the adaptive disparity estimation algorithm [5].

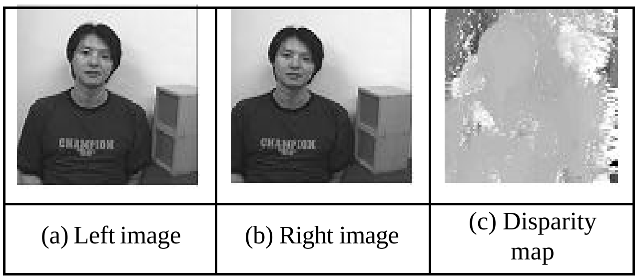

Fig. 6 Captured 2-view images and its disparity map

Here, the forward and reverse disparity maps are generated from the captured images and bi-directional disparity map is produced from two disparity maps. Then original left image and this bi-directional disparity map are 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 are used for generation of the 8-view images using the IVR scheme in the client side.

Figure 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 left top of the simulator screen.

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

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

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 shows 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.

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

$$PSNR = 20\log_{10}\left[\frac{255}{RMSE}\right] \quad (4)$$

And number of views, frames per second and delay time per frame are displayed at the right top of the simulator screen. In addition, the generated intermediate 16-view images are displayed in the lower part of the simulator screen. Because the IEEE 1394 adapter can transmit 50MB/sec per a port and up to 200MB/sec per 4 ports, there is no problem in transmission of the camera data. The computer system has three kinds of PCI buses which bandwidths are 132 MB/sec, 800 MB/sec, 1056 MB/sec, respectively. But the IEEE 1394a adapter and the graphic adapter can support only PCI bus, not PCI-X, so that the data rate is limited to 132 MB/sec for a PCI bus. So, in the proposed system, all PCI buses work as the general PCI devices. The graphic adapter of *Enseo MVP Pro ZX* has a PCI bus and 4 video-output ports as interface, so that it can support 4 display outputs. The graphic adapter used in this paper can support 8MB memory per port and 32MB memory per an adapter. From the experimental results, it is found that the display adapter can support 8-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 rate and number of views are mutually interrelated and can be easily controlled in the proposed system by using the developed software program so that, a lot of flexibility in design of the proposed multi-view 3D display system is 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 1600SW*. 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.

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

From these experimental results a possibility of implementing a new real-time web-based remote multi-view 3D imaging communication system using the proposed scheme is finally suggested.

4. Conclusions

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. From 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 15fps. This feasibility test for the proposed system also suggests a possibility for implementation of the web-based real-time multi-view 3D image communication system.

5. Acknowledgements

This research has been conducted by the ITRC project of the Korea Ministry of Information and Communication and the NRL project of the Korea Ministry of Science and Technology.

6. References

[1] M. Ebroul Izquierdo, Signal Processing 231 (1998) 5-7.

[2] K. H. Bae, J. J Kim, and E. S. Kim, *Optical Engineering* 42 (1998) 1778-1788.

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