28th Picture Coding Symposium, PCS2010, December 8-10, 2010, Nagoya, Japan

INFLUENCES OF FRAME DELAY AND PACKET LOSS BETWEEN LEFT AND RIGHT FRAMES IN STEREOSCOPIC VIDEO COMMUNICATIONS

Shuliang Lin, Yuichiro Sawa, Norishige Fukushima, Yutaka Ishibashi

Graduate School of Engineering, Nagoya Institute of Technology

ABSTRACT

This paper analyzes the influences of frame delay and packet loss on stereoscopic vision when stereoscopic video is transferred over a IP network. We employ live action videos which are transferred to a head-mount-display (HMD) and do the assessment on stereoscopic perception. As a result, we found that speed and movement direction of the attention object play a great role on the deterioration when frame delay and packet loss exist.

Index Terms— stereo video, network delay, packet loss, subjective assessment, inter-media synchronization

1. INTRODUCTION

Stereoscopic video communication allows a user to perceive perspective depth in a viewed scene remotely. But when this transmission happens on a network without QoS (Quality of Service) guarantee, such as the Internet, the quality of stereoscopic video can be easily deteriorated by delay or packet loss. Network delay brings disorder to the interval between left and right frames. Moreover, in the case when stereoscopic video is transferred separately, frame delay may damage the media synchronization and cause deterioration to stereoscopic vision. In the meantime, packet loss also influences stereoscopic perception by deferring the output times of both left and right frames. Yui *et al*. investigated the influence of frame delay on moving CG (Computer Graphics) objects in a virtual environment [1]. But they didn't study on the influence of packet loss, and they only used CG objects with simple movement without considering live action objects with combined complex movement. Thus using QoS controls to compensate these deteriorations becomes necessary, in order to implement QoS controls in the future, we investigate the influences on stereoscopic perception when live video is transferred in real time over an IP network. Mainly this paper only concerns the network factors on stereoscopic video communication.

In the following section, we outline the experimental environment we performed. Section 3 explains the assessment method. We present the results of the experiment in Section 4, and Section 5 concludes this paper.

2. EXPERIMENT ENVIRONMENT

In this section, we describe the details of our experiment. In the following subsections, we explain the experimental system and the stereoscopic videos we used.

2.1. Experimental system

In the experimental system, we employ an network emulator (NIST Net [3]) between the sending terminal and the receiving terminal, and they are connected with Ethernet cables. On the side of receiving terminal, the stereoscopic video is showed to each subject by an opaque type head-mount-display (HMD). This HMD is made by Vuzix and the type is iWear 920. The HMD has a maximum resolution at 640-by-480 with a update rate at 30 frames per second. In the case of frame delay, according to the left frame, the right frame is added with regular delay bilaterally. The left frames and right frames of the stereoscopic video are separately divided into fixed-length UDP datagrams to be sended. If the UDP datagrams of one frame do not arrive on the side of receiving terminal, this frame will be regarded as a frame loss. When frame missing happens, the latest frame will be outputted instead. In the case of packet loss, we use the network emulator to set the loss rate from 0 percent to 4.5 percent with intervals of 0.5 percent.

2.2. Stereoscopic videos

Five stereoscopic videos we used can be downloaded freely from web [2]. Their snapshots are shown in Fig. 1. Video 1 shows a trail view of a car and three oncoming cars. Video 2 shows a water stream pouring to hands. Video 3 shows a girl pouring wine to a champagne tower. Video 4 shows a performance by two clowns. Video 5 shows a lady training her dog. Length of these videos are 7.8 seconds (videos 1 and 2), 8.3 seconds (videos 3 and 4), and 5.0 seconds (video 5). Each one of these videos shows different objects having several movement speeds, and directions at the same time. Therefore, influences of frame delay and packet loss may be different depending on which object is paying attention. Thus we specify some objects for the subjects to watch, and call them Attention Object (AO). Table 1 shows AOs of each stereoscopic video.

978-1-4244-7133-1/10/$26.00 ©2010 IEEE

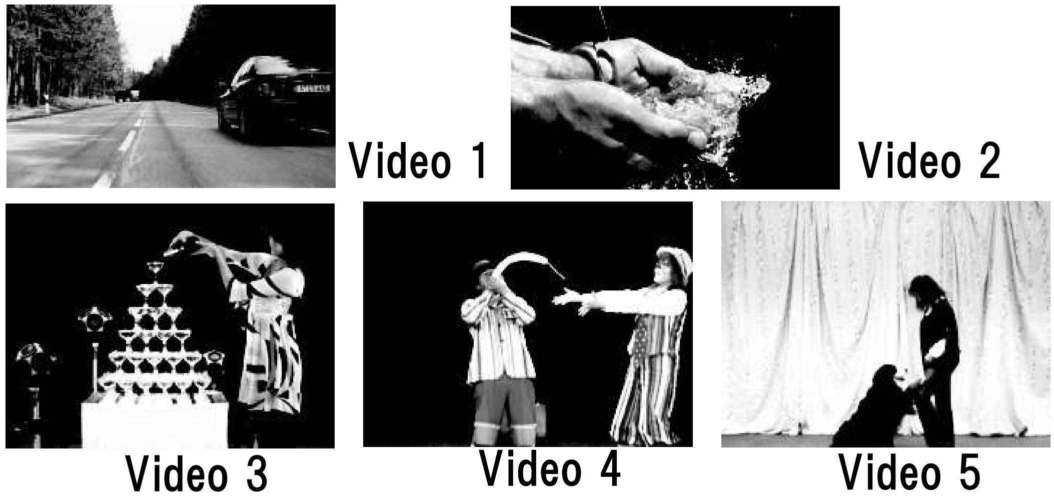

Fig. 1. Snapshots of videos

Table 1. Attention objects videos

Video 1	oncoming cars(1), front car(2)
Video 2	water spray(1), flowing water(2)
Video 3	woman pouring champagne(2)
Video 4	balloon(1), right clown's hands(2)
Video 5	turning dog(1), woman's stepping(2)

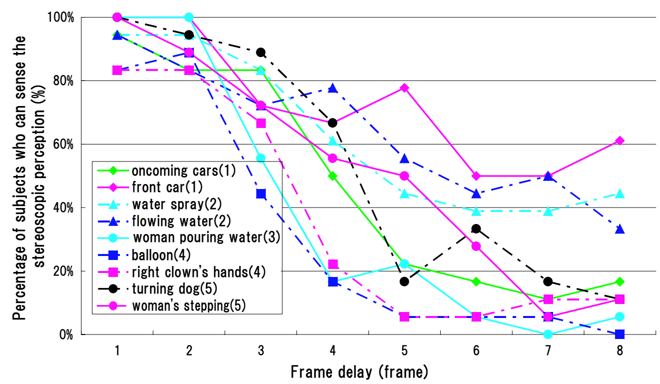

Fig. 2. Percentage of subjects who can sense stereoscopic perception under frame delay

3. ASSESSMENT METHOD

In the experiment, the subjects should see a stereoscopic video played repeatedly for 15 seconds, and then do the subjective assessment of whether they can sense the stereoscopic perception of the AO. Before the experiment, we show to the subjects the stereoscopic video without frame delay or packet loss and also the video with frame delay or packet loss.

In the subjective assessment, the subjects evaluate 9 AOs in random order. The frame delay is changed from 1 frame to 8 frames, and the number of lost frames per second is changed from 1 frame to 30 frames. The subjects are 30 men and women aged from 21 to 28 in the frame delay experiment, and 20 men and women aged from 21 to 25 in the packet loss experiment.

4. EXPERIMENTAL RESULTS

In this section, we show the results of frame delay and packet loss.

4.1. Frame Delay

Regarding the results of frame delay experiment, Fig. 2 shows the percentage of the subjects who can sense the stereoscopic perception. Fig. 3 depicts the average size of motion vector of attention objects of vertical movement(y-axis) which is most notable on the deterioration of stereo perception for human beside other two movement(x-axis and z-axis) or disparity. Fig. 4 shows the histograms of each notable point on the AO, and its horizontal axis represents the brightness quantized by 6-bit and normalized maximum frequency of 100.

In Fig. 2, when the frame delay is 2 frames or less, the percentage is above 80 percent. When the frame delay is larger than 2 frames, the percentage decreases sharply while the frame delay goes on increasing. Thus, we need to keep the frame delay under 2 frames in order to watch video with enough stereoscopic perception.

In Fig. 2, we can classify them into three groups: decreasing sharply; decreasing moderately; decreasing in degree which is between the previous two groups, and we regard this kind of group as middle group. The percentages of woman pouring champagne, balloon, and right clown's hands

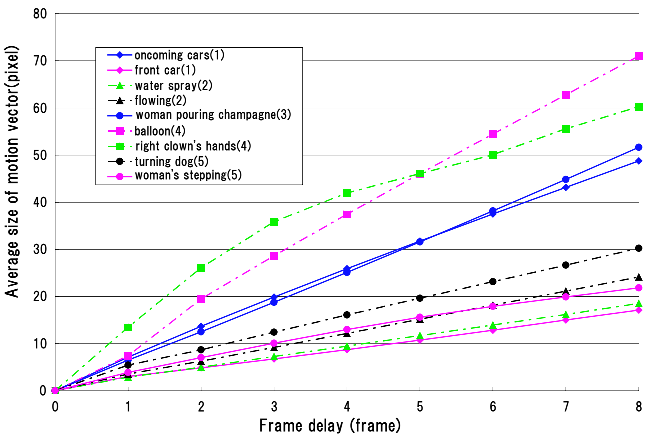

Fig. 3. Average size of motion vector (y-axis) between reference frame and delayed frame

all decrease sharply when the frame delay is more than 2 frames. It can be noticed that all the motion vectors of these AOs have large values in Fig. 3. This means that the large gap of corresponding points between right frame and left frame caused by large frame delay deteriorates the stereoscopic perception. In Fig. 4, we notice that the contrasts of AO and background in the cases of balloon and right clown's hands are high. Therefore this can be inferred that the deterioration of stereoscopic perception can be sensed more easily under larger brightness contrast of AO and background.

In Fig. 2, the percentages of oncoming car, water spray, and flowing water decrease moderately. We can find that their motion vectors in Fig. 3 are small. In this case, the gap between corresponding points on left frame and right frame is little, hence the influence on the stereoscopic vision is small. Oncoming car, turning dog, and woman's stepping belong to the middle group. In Fig. 2, the percentages of the middle group decrease more moderately than the decreasing sharply group. When the frame delay is 8 frames, they are similar to the sharply group's.

4.2. Packet Loss Rate

Figure 5 shows the percentage of subjects who can sense the stereoscopic perception under frame loss. We can see that along with the frame loss rate's increasing the percentages of woman pouring champagne, balloon and right clown's hands decrease more largely compared to other percentages of AOs. The reason is that the gap of corresponding points between left frame and right frame becomes larger while the speed of AO's y-axis movement goes on increasing. Also in the case of front car, the speed of AO's y-axis movement is relatively slow and its percentage line declines more slowly than others consequently. The rest of AOs' percentages decrease in degree between the previous two's.

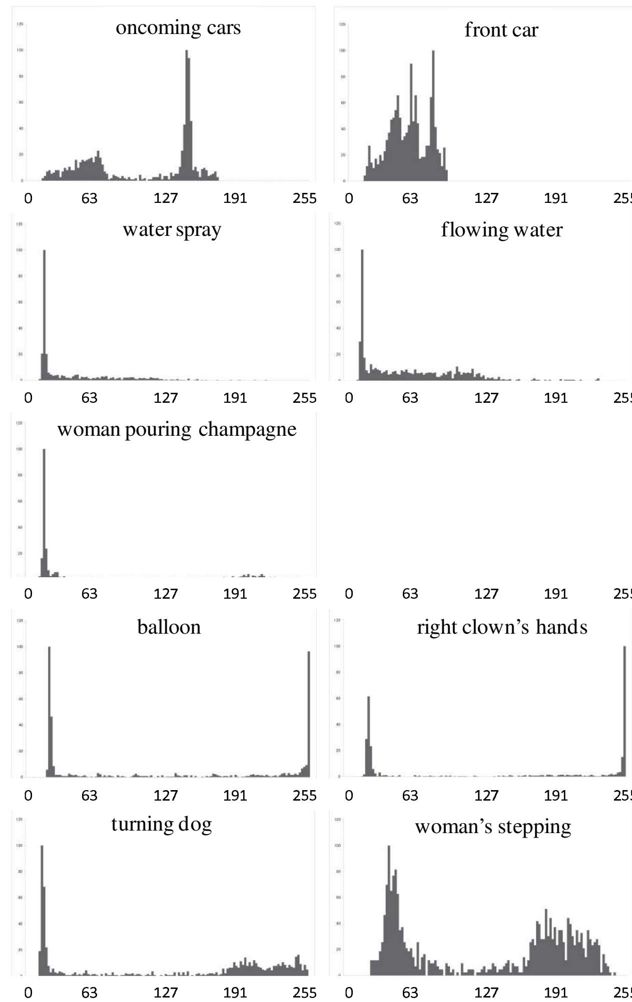

Fig. 4. Histograms of attention objects

5. CONCLUSIONS

In this paper, we investigated influences of frame delay and packet loss on stereoscopic video communications. As a result, we revealed that the range of frame delay when people can see video stereoscopically is 2 frames or less. Also we found that the gap between left frame and right frame can be perceived more easily when the noticing object has high speed in y-axis movement in both frame delay experiment and packet loss experiment.

As the next step of our research, we will study the influence of delay jitter along with QoS controls that can compensate delay jitter to smooth the video stream properly.

Acknowledgment

This work was partly supported by the Grand-In-Aid for Scientific Research (C) of Japan Society for the Promotion of

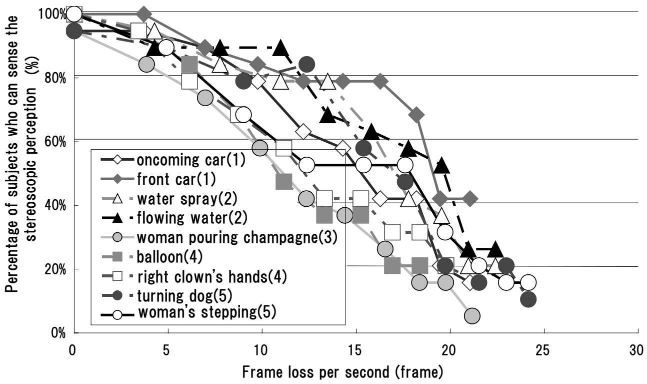

Fig. 5. Percentage of subjects who can sense the stereoscopic perception under frame delay

Science under Grant 22560368.

6. REFERENCES

[1] Y. Yui, H. Takahashi, M. Nakajima, "Effect of Frame Delay on Stereoscopic Vision Using Moving Pictures," in *Proc. IWAIT2004*, pp. 103–107, Jan. 2004.

[2] http://sp.cs.tut.fi/mobile3dtv/stereo-video/.

[3] M. Carson, D. Santay, "NIST Net: a Linux-based network emulation tool," in *ACM SIGCOMM*. Jul, 2003, vol. 33, no. 3, pp. 111–126.