Assessment of User Behavior and QoE in Multi-View Video and Audio IP Transmission

Erick Jimenez Rodriguez† Toshiro Nunome† Shuji Tasaka†
†Graduate School of Engineering, Nagoya Institute of Technology, Japan
Email: †{erickjim,nunome,tasaka}@inl.nitech.ac.jp

Abstract—This paper investigates QoE (Quality of Experience) of MVV (Multi-View Video) and audio transmission over IP networks. Unlike previous works that mainly discuss MVV transmission from aspects of video codecs, we study MVV and audio transmission under various IP traffic and delay conditions by experiment. We compare two schemes: a scheme that the user watches a single viewpoint, and the one that he/she can choose one viewpoint from many viewpoints. At the same time, we perform a user study which helps us to understand the user behavior in the MVV system under those conditions. As a result, we show that high QoE can be achieved when the user can change the viewpoint with low load traffic and delay.

I. INTRODUCTION

Television has a long history and keeps the position of the most important visual information system. It has realized the human’s dream of seeing a distant world in real-time. However, conventional TV shows us the same scene even if we move our viewpoints in front of the display. If it is quite different from what we experience in the real world. In TV, the users can get only a limited view of a real 3D world they want to see. The view is determined by the users but by a camera placed in the 3D world. Although many important technologies have been developed, this function of TV has never changed.

Because of this limitation in its functionality, MVV (Multi-View Video) [1], where the user can choose one video from multiple videos of the same event, has been under development. In addition, Free Viewpoint Television (FTV) [2] and 3DTV [3], which make use of MVV as their base system, have also been under investigation. In recent years, these technologies have been attracting owing to their enhanced viewing experience.

There are many challenges for implementation of MVV systems. One of these challenges is the way how a large amount of data should be streamed on the network with limited capacity. Because of this reason, there are several works which focus on compression algorithms for MVV (e.g., [4] and [5]).

As for the method to measure the performance of an MVV system, most of the related works employ PSNR (Peak Signal to Noise Ratio), which measures spatial quality of video, and the throughput. On the other hand, there has been relatively few researches on QoE (Quality of Experience) [6] assessment with MVV systems (e.g., [7] and [8]).

References [7] and [8] have performed a user study of interactive MVV systems. These references have assessed the effect of different features, such as viewpoint switching, frozen moment and viewpoint sweeping on the MVV system and the effect of the contents on the user’s behavior. However, they do not consider audio; in real applications, audio and MVV are transmitted together. In addition, they do not perform systematic QoE assessment when the delay and IP packet losses are present in the transmission.

In this paper, we perform several experiments where we examine the effect of the traffic and delay on the QoE and application-level QoS of MVV and audio IP transmission. We also study the effect of the ability of viewpoint change. We compare the QoE when the assessors watch only one viewpoint and that when they can choose one viewpoint from many ones.

In this paper, we refer to the former scheme as “Fixed view” and the latter as “Selective view”.

The rest of the paper is structured as follows. Section II outlines the system model. Section III discusses the conditions of the experiment we performed. We present the results of the experiment in Section IV, and Section V concludes this paper.

II. MVV SYSTEM MODEL

We will explain an outline of the MVV system in this paper. We assume one client and one server; they are connected to an IP network. Four cameras are connected to the server.

The server captures the video of each camera. At the same time, the audio is captured by a microphone. The server sends the audio and video of a viewpoint to the client by using UDP packets. The client receives these packets and outputs them.

The client can choose one viewpoint from the four cameras. The user notifies the MVV server of a desired viewpoint by using the viewpoint change interface. This request is sent to the MVV server by using a UDP packet. When the server receives this packet, it changes the viewpoint and start sending the video stream of the new viewpoint. The transmission lasts until when the server receives another request for viewpoint change or when the session ends.

III. EXPERIMENTAL CONDITIONS

A. Experimental system

Figure 1 shows the network topology used in the experiment. In this topology, MS is the server of the MVV application, and MR is the client. On the other hand, LS is the server of the background traffic, and LR is the client. Both router 1 and router 2 are Riverstone’s RS3000. At the same time, NN, which is a PC, is laid out between the routers. NN adds delay by using NISTNET [9] for data going through routers 1 and 2. By adding this delay, we can see the effect of delay on the QoE in the MVV system.
The average watching time on each camera is defined as the total time in seconds when a camera’s video is displayed.

D. QoE parameters

In this paper, we express QoE in terms of the interval scale, which is referred to as the psychological scale [10]. The interval scale can be calculated by one of the psychometric methods [11]. For the calculation of the interval scale, this paper adopts the method of successive categories, which is composed of two steps: the rating scale method and the law of categorical judgment. The rating scale method specifies how the subjective measurement is made on stimuli, which are audio-video streams output at the receiver in our case: an assessor classifies the stimuli into a certain number of categories (e.g., five) each assigned an integer (typically 5 through 1 in order of highly perceived quality). Since the law of categorical judgment is based on several assumptions, we have to confirm the goodness of fit for the obtained scale. For a test of goodness of fit, we conduct Mosteller’s test [12]. Once the goodness of fit has been confirmed, we use the interval scale as the QoE parameter, which is therefore called the psychological scale.

Every time the assessor sees an audio-video stream, he evaluates it according to the following criteria:

- Smoothness of the video
- Synchronization of the video and audio
- Quickness of the viewpoint change
- Overall evaluation

Each criterion is evaluated to be one of five levels between 1 (the worst case) and 5 (the best case). With the “Fixed view” scheme, all the criteria except “Quickness of the viewpoint change” are evaluated because the user does not change the viewpoint in this scheme. All the questions about the criteria are written in Japanese.

After one audio-video stream of each scheme has been shown, the user evaluates an additional criterion in order to express his opinion on which scheme was better under a given condition of delay and traffic. We refer to this criteria as “Fixed View vs Selective View”.

IV. ASSESSMENT RESULTS

In this section, we will present the experimental results of the application-level QoS assessment, the user’s behavior, and the QoE assessment.

A. Application-level QoS

Figures 3 through 5 show measured values of application-level QoS parameters as a function of the load traffic for the three values of the additional delay at NN: 0 ms, 100 ms, and 200 ms.

Figure 3 shows the viewpoint change delay for video with the “Selective view” scheme. We notice in this figure that for all the additional delay values, the viewpoint change delay increases as the UDP load traffic increases. This is because there is not enough bandwidth to send both UDP load traffic and the audio-video of our MVV application under the heavily loaded condition, and then the packets will be delayed or discarded.

Figures 4 and 5 show the MU loss ratio for audio and video, respectively, of the “Selective View” scheme versus the amount of load traffic.

We can see in Figures 4 and 5 that both audio and video have a similar tendency. As the traffic increases, the MU loss ratio also increases.

We have also assessed the MU loss ratio of the “Fixed view” scheme. As a result, we have noticed that the MU loss ratio of the “Fixed view” scheme is almost the same as that with “Selective view” scheme.
B. User study results

Figure 6 shows the average number of viewpoint changes versus the amount of the load traffic. We can see in this figure that the user changed the viewpoint more frequently when the traffic is light. As we have found in Figure 3, the user must wait to watch a new viewpoint when the network is heavily loaded. As this waiting time increases, the user becomes reluctant to change the viewpoint. For this reason, the traffic may affect the user’s behavior.

Figure 7 shows the average time when the users watched the video of each camera without additional delay. Here we can see that the users watched cameras 1 and 2 for longer time than the time when they watched cameras 3 and 4. This is because camera 1 is the default camera; the users always see camera 1 at the beginning, and in some cases they watch this camera for a longer time. As camera 2 is the next to camera 1, the users may want to change it to the following camera. For this reason, as the camera gets farther from the default camera, the user will watch the viewpoint for less time.

The content can affect the viewpoint change feature. We consider that the “exciting moment” of our content is the moment when the doll’s head is out of the viewpoint’s focus. At these moments, the user may be interested in changing the viewpoint to where the doll’s head is.

Since the user did not know the position of each camera exactly during the evaluation, he randomly changed the viewpoint regardless of where the doll’s head was. Because of this reason, whether the user knows the position of each camera or not affects the user’s behavior of our MVV application. In addition, the user changes the camera in turn until finding the viewpoint that he was looking for. If we improve the interface so that the user knows the camera position, we expect to improve the QoE; assessment of the effect of the interface on the user’s behavior is our future study.

C. QoE assessment results

We have found that the Mosteller’s test with a significance level of 0.01 cannot reject the hypothesis that the observed value equals the calculated one. Once the goodness of fit has been confirmed, we use the interval scale as the QoE parameter, which is therefore called the psychological scale. By doing this, we obtained the boundary for each category. The upper limits from Category 1 (C1) to Category 4 (C4) are shown in Table II.

In Table II, we notice that the width of one category is not uniform for each criterion. This means that the translation by the law of categorical judgment is mandatory.

Figure 8 shows the psychological scale value for the criterion of “Quickness of the viewpoint change” with the “Selective view” scheme. Here we can see that not only the UDP load traffic but also the delay affects the user’s evaluation. As both UDP load traffic and the delay begin to increase, the QoE for quickness of the viewpoint change also starts to decrease.

Figures 9 and 10 show the psychological scale value for the criterion of “Overall evaluation” with “Fixed view” and that with “Selective view”, respectively. Figure 11 shows the psychological scale value for the criterion of “Fixed view vs Selective view”.

We can see in Figures 9 and 10 that both schemes have a similar tendency; as the UDP load traffic increases, the overall
quality decreases.
In Figures 9, 10 and 11, we notice that the user seems to be more satisfied with “Selective view” only when both load traffic and delay are low. As the traffic and delay gradually increases, the QoE will gradually decrease until the point where the user is equally satisfied with the two schemes.

V. CONCLUSIONS

We made experiments on MVV and audio transmission under various IP traffic and delay conditions. We compared the two schemes: “Fixed view” and “Selective view”. We assessed the effects of the IP traffic and the delay on QoE for the two schemes.

From the application-level QoS evaluation results, we saw that the ability of viewpoint change does not affect the application-level QoS of our MVV application. In addition, as the UDP load traffic begins to increase, viewpoint change delay also increases.

From the user study results of our subjective evaluation, we observed that users tend to change the viewpoint frequently in light traffic and low delay. Also, they tend to watch the default camera and near cameras for longer time.

From the QoE evaluation results, we found that the user prefers the “Selective view” scheme in low load traffic and low delay. However, as the traffic and delay gradually increases, the QoE will gradually decrease until the point where the user is equally satisfied with the two schemes.

We noticed that QoE with the “Selective view” scheme was not good as excepted. The user only prefers the scheme under the limited condition. For this reason, we will continue to study how the application should be improved to expect a better QoE.

As mentioned in the user study results, the interface used to change the viewpoints can affect the usage of our MVV application and the QoE. For this reason, we will study the effect of the interface in our MVV application, where the user can change the viewpoint using the camera direction rather than using the camera number.

REFERENCES