Traffic Reduction for Multiple Users in Multi-view Video Streaming

Takuya Fujihashi
Faculty of Informatics
Shizuoka Univ.
Hamamatsu, Japan
fujihashi@aurum.cs.inf.shizuoka.ac.jp

Ziyuan Pan
Graduate School of Informatics
Shizuoka Univ.
Hamamatsu, Japan
pan@aurum.cs.inf.shizuoka.ac.jp

Takashi Watanabe
Graduate School of Informatics
Shizuoka Univ.
Hamamatsu, Japan
watanabe@inf.shizuoka.ac.jp

Abstract—Multi-view video consists of multiple video sequences captured simultaneously from different angles by multiple closely spaced cameras. It enables the users to freely change their viewpoints by playing different video sequences. Transmission of multi-view video requires more bandwidth than conventional multimedia. To reduce the bandwidth, UDMVT (User Dependent Multi-view Video Transmission) based on MVC (Multi-view Video Coding) has been proposed for single user. In UDMVT, for multiple users the same frames are encoded into different versions for each user, which increases the redundant transmission. For this problem, this paper proposes UMSM (User dependent Multi-view video Streaming for Multi-users). UMSM possesses two characteristics. The first characteristic is that the overlapped frames that are required by multiple users are transmitted only once using the multicast to avoid unnecessary duplication of transmission. The second characteristic is that a time lag of the video request by multiple users is adjusted to coincide with the next request. Simulation results using benchmark test sequences provided by MERL show that UMSM decreases the transmission bit-rate 55.3% on average for 5 users watching the same multi-view video as compared with UDMVT.

Keywords-Multi-view Video, Multi-users, Realtime Streaming, Multicast, Traffic Reduction

I. INTRODUCTION

Multi-view video taken by multiple cameras from the same scene is possible because the camera is smaller and lower priced. It enables users to freely change their viewpoints whenever they want to. There are several applications for multi-view video, such as Free Viewpoint TV (FTV) [1], [2] and remote medical surgery. Furthermore, improvement of transmission technology makes multi-view video watched by multiple users possible.

Since multi-view video consists of multiple video sequences, the transmission bit-rate of multi-view video is several times larger than traditional multimedia, which imposes more bandwidth requirements. To reduce multi-view video traffic, compression and transport technology are important. As the previous research on the efficient transmission of multi-view video shows, there are two kinds of methods.

The first method is that all video sequences are sent to each client to avoid the switching delay. Typical technique includes Multi-view Video Coding (MVC). However, even with the MVC, transmission bit-rates for multi-view video are still high: about 5Mbps for 704 × 480, 30fps, and 8 camera sequences with MVC encoding [3].

The second method is to transmit only the frames requested by the user [4], [5]. UDMVT (User Dependent Multi-view Video Transmission) [5] is one of these techniques. UDMVT only sends the frames depending on the user’s motion, which is fed back periodically. However, UDMVT does not support the multiple users well. When there are many users watching the same video, many duplicate frames are encoded and transmitted. Therefore, the traffic will increase with the increase in the number of users.

In this paper, we discuss the area of overlap frame for multiple users and propose UMSM (User dependent Multi-view video Streaming for Multi-users) to reduce the transmission of overlap frames between users. In UMSM, feedback from users is used to calculate the overlap of frames. The overlap frames are encoded once for all the users who need them and then the encoded frames are transmitted to all these users with multicast. The server also adjusts the feedback time to close the gap in the request of users. Therefore, UMSM prevents duplicate transmission of the same frames for multiple users to reduce transmission bit-rates for multi-view video. Furthermore, we discuss a trade-off between transmission count and transmission bit-rate by changing the threshold of the size of overlap frames. Performance evaluation shows that UMSM reduces the transmission bit-rates of multi-view video transmission when there are multiple users watching it.

II. RELATED WORKS

A. MVC

MVC is issued as an amendment to H.264/MPEG-4 AVC [6]. The key of MVC combines the temporal prediction and inter-view prediction together.

Statistical evaluations in [7], [8] show significant compression gains and decrease of transmission bit-rate. However, the prediction structure of MVC makes the views depend on each other. In order to display the multi-view video correctly, the frames that are displayed frames and the frames on which they depend must be received first. It will cause more unnecessary transmission and delay as the displayed views would be far away from the reference view.
B. UDMVT

MVC has a high transmission bit-rate because it transmits every view and many frames are unnecessary. UDMVT [5] is transmission technology that analyzes the user’s motion to prevent the transmission of redundant frames.

UDMVT focuses on the successive motion switching model in which the user is able to switch from current views to neighboring views. In other words, if the multi-view video contains the views (1, 2, ... M), for any view j the user is just able to switch from j to the view j’, where max(1, j − 1) ≤ j’ ≤ min(j + 1, M). In the successive motion model, frames that should be displayed when the user starts to switch to the next view are decided by both the frame rate (frame/sec) of the multi-view video and the switching speed (view/sec) of the user. Let k be the floor of the frame rate divided by switching speed: k = \lfloor f/s \rfloor in which f denotes frame rate while s denotes the switching speed of user.

Therefore, if the user periodically feeds back the three-tuples N(p, f, s), the server can predict an area in which frames may be displayed in the next period of time as shown by Fig.1. p is the initial position F_{i_o,j_0} which is the frame of view j_0 at time i_0. f is the frame rate while s is the switching speed. Although these three-tuples cannot exactly predict all the frames in the display path, they are able to predict a triangle area in which the frames can be displayed in the next period of time. The main idea of UDMVT is that only the frames in the area (called Potential Frame) are encoded and transmitted and the frames out of the area (called Redundant Frames) are ignored. Therefore, the UDMVT can reduce the transmission bit-rate for the transmission of multi-view video.

UDMVT transmits only the necessary frames for the user according to the periodic feedback from the user. As the number of views increases, UDMVT reduces more transmission bit-rate than MVC. However, UDMVT does not support the multiple users. The frames encoded for one user cannot be used by other users. When there are many users watching the video, many duplicate frames are encoded and transmitted. Therefore, with the increase in users, the transmission bit-rate becomes higher and higher.

III. MULTI-VIEW VIDEO WITH MULTI-USERS

A. Problem of Multiple Users on UDMVT

When there are multiple users watching the same multi-view video, each user has his own potential frame area (PFA). As shown in Fig.2, it is assumed that there are two users watching the same multi-view video at the same time. User 1 is watching view 2. After 1 or 2 frames, user 2 is watching view 4. The number of these shifting frames between users is called offset (frames). Fig.2 (a) shows the PFAs of user 1 and user 2 when k = 1 and offset = 1 while PFAs with the k = 2 and offset = 2 are shown in Fig.2 (b).

When offset is not 0 among multiple users, it is hard for the server to encode the common frames for these users at the same time. As mentioned in Section II, the frames encoded for one user cannot be used by other users. In the PFA, the first frame is encoded as one I-frame. Other frames in the same PFA are predicted directly or indirectly from this I-frame. As in Fig.2 (a), frame 1 and 2 are both encoded into I-frame for user 1 and user 2, respectively. The frames in deep color are the frames in the overlap of two PFAs. In UDMVT, each frame in the overlap is encoded into two versions: one is predicted from frame 1 and the other is predicted from frame 2. These two versions of each frame are transmitted to user 1 and user 2, respectively. These frames in the overlap are called Overlap Frames (OFs). Therefore, as the number of users increases, the traffic of UDMVT will also increase.

B. UMSM

If the frames in the overlap, which is called Overlap Frame Area (OFA), can be encoded just once and multicast to all the users, it can reduce the duplicate encoding and transmission. It is helpful in reducing the traffic of multi-view video transmission. Therefore, when the server receives the feedbacks from users, it will calculate the OFA between the PFAs.

Therefore, let N_{i,o}^{x}(p^x, f, s^x) be the information fed back by the user x. p^x is the start frame F_{i_0,j_0}^x (frame in the view j_0 at the time i_0) of user x. f is the frame rate while s^x refers to the switching speed of user x. As the number of the views is M, \mathbb{R}^x(t) is the set of frames that can be displayed by user x at time t when started from F_{i_0,j_0}^x. F_{i,j} denotes the frame of view j at time instant i. F_{i,j'} \in \mathbb{R}^x(t), in which:

\[ i = i_0 + \lfloor f \times t \rfloor \]

\[ j' \in \text{max}(1, j_0 - \lfloor s^x \times t \rfloor), \text{min}(j_0 + \lfloor s^x \times t \rfloor, M) \]

Thus, the PFA of user x is expressed as Eq. (1) when the

---

Figure 1. Feedback scheme for UDMVT.

Figure 2. The range of transmission frames with k. (a) k = 1 and offset = 1. (b) k = 2 and offset = 2.
length of the PFA is \( L \).

\[
PFA^x = \bigcup_{t=0}^{L-1} R^x(t) \tag{1}
\]

Overlap Frames are the frames in the overlap of the PFAs. Therefore, the OFA of user \( x_1 \) and user \( x_2 \) can be achieved by:

\[
OFA(x_1, x_2) = PFA^{x_1} \cap PFA^{x_2}. \tag{2}
\]

After calculating the OFA, the OFs are encoded for all users and transmitted with multicast in UMSM.

C. Adjustment of Offset

Each user should feedback the information of \( N^x(p^x, f, s^x) \) periodically at the end of the PFA for encoding and transmission of the next PFA. In order to continuously play the video without interruption, the feedback should be sent shortly before it is displayed to the end of current PFA because of the delays. The end of current PFA is related to the delay between starting a feedback and the time when the received frames are ready to be displayed. This delay contains four components: feedback delay \( t_1 \), encoding delay \( t_2 \), transmission delay \( t_3 \) and decoding delay \( t_4 \). The time \( t_1 + t_2 + t_3 + t_4 \) is the sum of delay between requesting PFA and obtaining the displayable frames. When the frames are received, they are displayed according to the user switching. Let \( \tau_1 \) be the feedback period of user 1 and \( \tau_2 \) be the feedback period of user 2.

When displaying multi-view video for multiple users, the time of receiving the feedback from different users is different. This feedback time difference is the offset among multiple users. If this offset remains, frames may be encoded separately between the users. Frames for user 1 may be have already been encoded when the feedback of user 2 arrives. Therefore, UMSM may not be used to reduce the traffic. To adjust offset between users to an acceptable range, UMSM temporarily changes the feedback time for each user at the server.

Fig.3 shows the time sequence chart of the UMSM. It is assumed that view 9 is watched by user 1 while user 2 begins to watch the multi-view video. View 5 is selected by user 2. The feedback is sent to the server when user 1’s initial position is frame 100 at view 9. After offset, user 2 sends a request to the server when user 2’s initial position is frame 150 at view 5. If there was already feedback from other users, server decreases the user’s feedback period by offset for the next feedback. Therefore, in the next PFA, user 1 switches to view 8 and user 2 continues to watch view 5. In order to continuously play the video, the \( N^{x_1}(p^{x_1}, f, s^{x_1}) \) should be detected and fed back after time \( \tau_1 \) after the previous feedback by user 1. For user 2, because the feedback time is reduced to \( \tau_1 - \) offset for the next feedback, next time user 2 will be fed back nearly at the same time as user 1. When only one user is watching the multi-view video, UMSM uses the UDMVT method.

At the next feedback, the initial position of user 1 is frame 200 at view 8 and user 2 is frame 200 at view 5. The server receives the feedback and finds the user’s initial position for next PFA. When the feedback period is different for each user, the server will synchronize the users’ feedback period to the smallest feedback period as the following equation: \( \tau = \min(\tau_1, \tau_2) \). Next, PFAs will be encoded and transmitted starting from frame 200 at view 8 and frame 200 at view 5 for user 1 and user 2, respectively. UMSM then calculates the OFA of user 1 and user 2. The OFs in the OFA and the PFs outside the OFA (called Special Frames) will be encoded. In UMSM, OFs in the OFA will be multicasted to multiple users and the Special Frames (SFs) will be unicasted to user 1 and user 2, respectively. In this way, UMSM can reduce duplicate encoding and transmission to decrease the transmission bit-rate.

D. Prediction Structure

For UMSM, in order to avoid duplicate transmission of OFs, OFs should be encoded once and should be used by all the users who need them. To achieve this, either the OFs are encoded independently or they are decoded by the PFs of any user. Therefore, in the UMSM, the anchor OF can be encoded in two ways under different conditions as follows:

First, the frame \( OF_{i_0,j_0} \) is selected as anchor frame as follows:

\[
i_0 = \min(i) \text{ where } F_{i,j} \in OFA. \tag{3}
\]

\[
j_0 = \min(j) \text{ where } F_{i_0,j} \in OFA. \tag{4}
\]

Then, \( OF_{i_0,j_0} \) is encoded into an I-frame or an SP-frame:

\[
OF_{i_0,j_0} = \begin{cases} 
1 - frame & \text{when } F_{i_0,j_0} \in PFA^x, x \neq x'. \\
SP-frame & \text{else}.
\end{cases} \tag{5}
\]

When the anchor frame of one \( PFA^x \) is in other \( PFA^{x'} \), which means the \( PFA^x \) is the OFA of x and \( x' \), the anchor frame of OFA should be encoded into an I-frame as shown.
by Fig.4 (b). Other OFs are predicted from this I-frame directly or indirectly. Therefore, when user $x$ obtains the frames of OFA, all the PFs needed by user $x$ can be decoded without any other PFs belonging exclusively to user $x'$. Otherwise, the anchor frame of OFA is encoded into an SP-frame. SP-frame [9], [10] based on H.264/MPEG-4 AVC is standardized in order to maintain low compression rate and network friendliness for switching video streams. The SP-frame can be reconstructed by using different predictors. In UMSM, SP-frame is predicted from all the anchor frames of PFAs that are overlapping as shown in Fig.4 (a). The SP-frame is predicted from the two I-frames of user 1 and user 2. Then, when user 1 and user 2 receive the SP-frame, they can decode it by using their own I-frame.

Other OFs are predicted from the anchor frame of OFA directly or indirectly. In order to maintain a low compression rate, the SFs can be predicted from the OFs according to the need of each user. Therefore, after receiving the OFs and SFs, each user can decode all the frames in the PFA without any frame from other users. The OFs can be encoded once and used by all the users who need them.

E. The Way to Transmit OFs and SFs

When increasing the number of users who are watching the same multi-view video, the OFs will be separated into many OFAs and the traffic of multi-view video will decrease. However, the count of transmissions will rise. If raising the transmission count, the server needs to perform many multicast sessions to transmit OFs. Thus, router’s storage overhead and processing cost increase [11]. Therefore, there is a trade-off between the traffic of multi-view video and the count of transmissions.

This section explains the way to transmit OFs and SFs. At first, we initialize the set of PFA and OFA for $N$ users($F_N$), recursively. $OFA^n$ means all the combinations of $n$ users’ OFAs. $U$ is the set of users and $N$ is the number of users. $PFA^x$ is the PFA of user $x$.

$$F_n = \begin{cases} \bigcup_{x \in U} PFA^x & \text{if } n = 1, \\ F_{n-1} \bigcup OFA^n & \text{else.} \end{cases} \quad (6)$$

Next, we define the recursive function $f(x)$, and then we perform the $f(N)$ to determine the way to transmit each set of frames in $F_N$. We define the threshold of the size of OFA ($\bar{S}$) according to the multicast need of multiple users in order to support the trade-off between transmission and traffic. When the size of overlap frames $S(OFA(x_1, \ldots, x_N))$, where $OFA(x_1, \ldots, x_N) \in F_N$ and $S(x)$ is the size of frame area $x$, is larger than $\bar{S}$, UMSM multicasts the OFs to multiple users 1 to $n$ because the OFAs are large enough for multicasting.

Otherwise, the OFs do not multicast to users because the size of OFA is too small to multicast. Multicast of the small sized OFs will increase the count of transmissions and will not decrease the traffic. In this case, UMSM is trying to multicast the large OFs to multiple users instead of sacrificing the traffic of multi-view video. By applying this method to all OFA, each OF determines the multicast destination. Finally, the SFs in $F_N$ are unicasted to each user.

The details of algorithm $f(x)$ are described as Algorithm 1. $T$ is the set of transmission frames.

**Algorithm 1 Recursion $f(x)$**

**Require:** $F_N, U, T, S$

1. If the value of $x$ is greater than 1 then
   2. While OFAs of $x$ users remain in $F_N$ do
      3. If the value of $x$ is greater than 2 then
         4. If the size of OFA is larger than $\bar{S}$ then
             5. Add the OFs to the set of transmission frame $T$
             6. Subtract the OFA from frame list $F_N$
         7. End if
         8. Remove the OFA from frame list $F_N$
      9. Else
         10. Add the OFs to the set of transmission frame $T$
         11. Subtract the OFA from frame list $F_N$
     12. End if
   13. End while
14. Recursion on $x - 1$
15. Else
16. Add the SFs to the set of transmission frame $T$
17. Terminated these procedure
18. End if

At first, we compare the value of $x$ with 1. When $x$ is greater than 1, $f(x)$ continues to be performed while $x$ users’ OFA remain in $F_N$. If any OFAs remain in $F_N$, this algorithm focuses on the value of $x$ to decide post-processing. When $x$ is greater than 2, UMSM compares the $x$ users’ OFA size with $\bar{S}$. The OFs decide to multicast and are added to $T$. After that OFA is removed from $F_N$, when the size of OFA is more than $\bar{S}$. Furthermore, the OFA subtract from $F_N$ because these OFs do not need to re-transmit. Otherwise, when the size of OFA is less than $\bar{S}$, OFA is removed from $F_N$ directly and focused on large OFA, including this OFA.

If the value of $x$ is less than 2, UMSM multicasts the OFs to multiple users forcibly because it is the largest OFA in $F_N$. When the $x$ users’ OFs do not exist in $F_N$, UMSM performs the function $f(x - 1)$, recursively. UMSM continues the above process until $x = 1$.

When $x = 1$, UMSM decides to unicast the SFs to each
user, and then the SFs add to \( T \). As a result, we define the count of transmission \( C(T) \) and the transmission bit-rate of multi-view video \( B(T) \) as follows:

\[
C(T) = |T| \\
B(T) = \sum_{f \in T} S(f)
\]  

(7)  
(8)

F. Multicast in Network Layer

UMSM is an application layer protocol for multi-view video and it uses the multicast protocol in the network layer. More exactly, UMSM transmits OFs by multicasting and SFs by unicasting.

There are two possible ways to apply the multicast to UMSM: 1) using a conventional multicast protocol such as MOSPF and 2) using the specialized multicast protocol for multi-view video transmission. For the latter approach, [12] proposed the extension of multicast protocol NICE [13] for multi-view video of multiple users. [12] shows that the protocol is effective for more than 40 users. However, we believe the users viewing the same video in real time should not exceed 40. Therefore, UMSM assumes the number of users will be less than 10 and it uses the conventional multicast protocol in the network layer.

IV. PERFORMANCE EVALUATION

The evaluation results have been obtained by using the multi-view video test sequence “ballroom” with 8 views and 640 × 480 resolution. This test sequence is provided by MERL [14]. The reference techniques are MVC and UDMVT.

Encoders implemented by the modified open source project JMVC [15] and JM 18.0 [16] were used to encode the multi-view video sequences. In each view, 250 frames were encoded with the frame rate 25 f/s. The length of the GOP was set as 8. Four values of the k were used: 1 to 4. The number of users was from 2 to 10. The length of feedback period was set as one GOP.

A. Overlap Rate

Overlap rate means the ratio of OFs to users’ PFs. 0% means PFs do not overlap at all between the users, so that UMSM will have the same performance as UDMVT. 100% means PFs is overlapping perfectly. This is an important factor for UMSM, because it determine whether the UMSM method can be applied effectively to multi-view video transmission for multiple users.

Fig.5 shows CDF (Cumulative Distribution Function) of overlap rate when the number of views is 8 and user’s values of k is 1 to 4. When there are 2 users, the percentage of 0% in overlap rate is 8.6% and the average of overlap rate is 38.4%. Then, with the increasing number of users, the average of overlap rate increases. The average of overlap rate is 58.5% in 3 users, 69.8% in 4 users, and 76.5% in 5 users. In addition, when there are 3 or more users, the percentage of 0% in overlap rate is 0%. This means, when the users switch views quickly or the number of users is larger, UMSM boosts the chance of reducing the traffic more than UDMVT does.

B. Traffic of Multi-view Video

Fig.6 shows the transmission bit-rate of each technique, where the 3 users watch the same multi-view video with varying values of k. Then, each user switches views randomly in 250 frames. This simulation is performed 10 times, and then the average of values is obtained. Also, we set the threshold of OFA’ size (\( S \)) is 0 (bit). In Fig.6, when the value of k is 1, UMSM reduces the transmission bit-rate of UDMVT about 40.2% and MVC about 51.9%, respectively. Further, when the value of k is 2, UMSM also decreases the transmission bit-rate of UDMVT about 34.5% and MVC about 64.0%. When the value of k is 1, which means users switch quickly, the increase of PFA will increase the transmission bit-rate. Therefore, the possibility of 3 users’ OFA increases. As a result, UMSM decreases the transmission bit-rate of multi-view video for multiple users because the OFs in the OFA are encoded and transmitted only once.
Then, as value of \( k \) increases, the frames need to be encoded and the transmission needs to decrease. Thus, the transmission bit-rates of UDMVT and UMSM are lower than MVC when the value of \( k \) is 1. However, as the area of each PFA becomes smaller, the OFA between 3 users also become smaller. Our proposal can reduce more transmission bit-rate than UDMVT only if there is any OFA between users.

Fig.7 shows the transmission bit-rate of UDMVT and UMSM where \( k \) is 1 with a varying number of users. Other simulation parameters are the same as above. As a result, UMSM decreases the transmission bit-rate of UDMVT about 48.8% when 4 users. Also, with 5 users, UMSM reduces about 55.3% more than UDMVT. Moreover, the results of the reducing ratio of 6 and more users are shown in Table. 1. As the number of users increases, the overlap of users’ PFA increases as shown in Fig.5. Therefore, with an increase in users, UMSM reduces more transmission bit-rate because UMSM transmits each OFA only once between multiple users.

C. Trade-off between Transmission and Traffic

Fig.8 shows the trade-off of transmission count and the transmission bit-rate with a variation in the size threshold of OFA (\( \tilde{S} \)). This simulation performs on 10 users, the value of \( k \) is 1, and other parameters are the same as in Section I V.B.

When the value of \( \tilde{S} \) is 0 (kbit), UMSM transmits the each OFA only once. Thus, the bit-rate of transmission is best; however, transmission count is worst. On the other hand, when the value of \( \tilde{S} \) is large enough, such as 1000 (kbits), transmission count is best. However, the transmission bit-rate increases. Increasing the transmission count means the server needs to build more transmission flows, place the router’s storage overhead and process the cost of routers.

As shown in the results, the optimized value of \( \tilde{S} \) is 350 (kbits). Then, when we decreased the number of users to 5 and 8, this value mostly did not change. However, when the users increased to 13, this value greatly decreased to about 100 (kbits). The reason is because that, as the number the users increases, each OF is separated into many small OFAs. Therefore, when the \( \tilde{S} \) is low, this value has a significant impact on the trade-off between transmission count and transmission bit-rate.

V. Conclusion

In this paper, we proposed a User dependent Multi-view video Streaming for Multi-users (UMSM) to reduce the encoding and transmission of overlap frames. At first, we discuss the range of overlap frames for multiple users. UMSM transmits the overlap frames to multiple users by multicasting and the special frames to each user by unicasting. Also, UMSM adjusts the user’s feedback time at the server in order to encode all the frames that are needed by multiple users. Furthermore, we discussed the trade-off between transmission and traffic in addition to changing the threshold size of overlap frame areas.

Simulation results of 2 to 10 users using test sequences provided by MERL show that UMSM achieves lower transmission bit-rate than UDMVT. For example, when the users switch to views randomly, UMSM achieves the transmission bit-rate 55.3% less on average than UDMVT on 5 users.

REFERENCES