A Discussion on Multi-view Video Streaming in Wireless Networks

Takuya Fujihashi, Ziyuan Pan and Takashi Watanabe

Faculty of Informatics, Shizuoka University
3-5-1 Johoku, Hamamatsu-Shi, Shizuoka, 432-8011, Japan
e-mail:{fujihashi, pan, watanabe}@aurum.cs.inf.shizuoka.ac.jp

ABSTRACT

Several techniques have been proposed such as SCC, MVC and UDMVT for transmission of multi-view videos which requires more bandwidth than conventional multimedia. However, there is no sufficient discussion on the streaming over wireless networks, where the latency and packet loss increase due to varying conditions. This paper discusses the multi-view video streaming on wireless networks. At first, problems of multi-view video streaming in wireless networks are addressed and existing techniques are compared to find loads of encoders and decoders and network traffic. Then, a framework with two-step encoding to reduce the latency and packet loss is presented. The primary encoder uses SCC or MVC depending on server capacity in wired networks, while the secondary encoder uses UDMVT in wireless networks for decoded stream primarily encoded. We evaluate SCC, MVC and UDMVT for the secondary encoder. Results show that UDMVT outperforms others regarding transmission bit-rate, latency and packet loss.

Keywords: Multi-view Video, Wireless Network, Framework, Two-step Encoding

1 Introduction

Multi-view videos taken by multiple cameras of the same scene from different positions and angles become feasible since small sized and less price cameras are available. Multi-view video allows the clients to freely change their viewpoints whenever they want to see the scene from any position. There are many applications for such multi-view video, such as remote security monitoring, remote medical surgery, undersea surveillance systems and Free Viewpoint TV(FTV) [1][2].

Fig. 1 shows multi-view video architecture. Video frames taken by several cameras are encoded by the encoder in the server. The encoded data is transmitted over the network and then decoded by the decoder to provide the views for clients. Since a multi-view video consists of multiple video sequences, the bit-rate of multi-view video is several times larger than traditional multimedia, which brings the significant increase in the bandwidth requirement. However, multi-view video taken different videos from the same scene, the views are highly correlated. To reduce multi-view video traffic and transmit the encoded data, compression and transport technology is important. The previous works on the efficient transmission of multi-view video are classified into two ways.

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

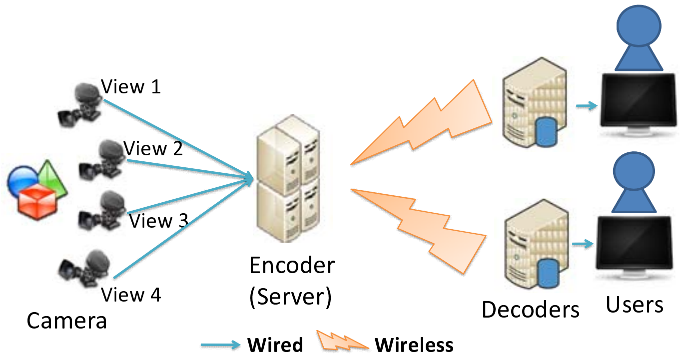

Figure 1: Multi-view video architecture.

The second way is transmit try only the frames requested by the client [4]–[6]. UDMVT (User Dependent Multi-view Video Transmission) [5][6] is one of these techniques. UDMVT sends frames depending on the client's motion which is fed back periodically.

Recently, video streaming over wireless networks become feasible because of the rapid progress of wireless technologies. However, the communication condition of wireless networks is varying in time. If we do not deal about that variation, the network latency and packet loss will increase. The increasing of packet loss causes the error propagation of frames in video streaming. Therefore, the image quality and interactivity of streaming will be deteriorated. In this paper, we investigate the problems that occur in wireless networks for multi-view video streaming, such as network latency, packet loss and limitation of battery life and bandwidth. Then, a framework with two-step encoding to reduce the latency and packet loss is presented. The primary encoder uses SCC or MVC depending on server capacity in wired networks, while the secondary encoder uses UDMVT in wireless networks for decoded stream primarily encoded. We evaluate the framework from a viewpoint of what coding technique should be used for secondary encoder in wireless networks by the network simulater, Qualnet. Finally, we discuss the evaluation results.

This paper is organized as follows: We introduce existing works related in Section 2. The problems of multi-view video streaming over wireless networks are described in Section 3. Based on Section 3, we present a framework to reduce the processing, latency and packet loss in Section 4. Section 5 presents the performance evaluation of presented framework compared to SCC, MVC and UDMVT by Qualnet. Finally, section 6 gives the conclusion.

2 Related Works

2.1 SCC(SimulCast Coding)

The Moving Picture Experts Group (MPEG) is widely used as the format of digital television signals. Encoding of video information is achieved by using two main techniques termed spatial and temporal compression in MPEG. Spatial compression involves analysis of a picture to determine redundant information within that picture while temporal compression is achieved by only encoding the difference between successive pictures.

In MPEG, there are three types of frames: I-frame, P-frame and B-frame. I-frame is an intra-coded picture and does not require other video frames to decode. P-frame contains the difference information from the preceding I- or P-frame. B-frame contains the difference information from the preceding and following I- or P-frame. The straight-forward solution for multi-view video encoding is simulcast encoding in which all video sequences are encoded independently using MPEG compression technology. However, simulcast encoded video still contains a large amount of inter-view redundant information.

2.2 MVC(Multi-view Video Coding)

A large amount of inter-view redundant information is still contained in simulcast encoded video. In order to remove the correlation between views, MVC is issued as an amendment to H.264/MPEG-4 AVC [7]. The key of MVC is combining the temporal prediction and interview prediction together. Each picture is able to predict from both neighboring picture temporally and the corresponding pictures in adjacent views. Statistical evaluations in [8][9] also show that significant compression gains and decrease of bit-rate can be expected from the combination of temporal prediction and interview prediction.

However, the prediction structure of MVC makes the views depend on each other. In order to display the multi-view video correctly, the frames which are displayed and the frames they depend on must be received at first. It will bring more unnecessary transmission and latency as the views should be displayed is far away from the reference view.

2.3 CDSS(Client Driven Selective Streaming)

To reduce the transmission bit-rate, [10] proposes the protocol that combines MVC and SVC (Scalable Video Coding) called CDSS (Client Driven Selective Streaming). In this protocol, the view that client needs to is decided by the predictor on client. Then, server encodes the view into two quality levels of base layer and enhancement layer. The base layer encodes the all views into lower bit-rate using MVC. The enhancement layer encodes the views which are selected by client in order to allow random access and improve the quality of base layer. However, the performance of this system depends on the Kalmanfilter-based predictor. If there are no prediction errors, the high-quality streams are displayed. However, the predictor is not hundred percent exactly. If the prediction is incorrect, only the base layer (low-quality) is displayed and it brings the bad client experience.

2.4 IMVS(Interactive Multi-view Video Streaming)

In [4], an algorithm of building optimized frame structure in order to reduce transmission traffic on IMVS (Interactive Multi-view Video Streaming) is proposed. This paper assumed Store & Playback which means the multi-view video is pre-encoded and stored in server. Then, the server transmitted the encoded frames whenever client requests it.

At first, this method initializes the frame structure to DSC (Distributed Source Coding)-frame which is proposed in [11]. DSC-frames have a good performance of the trade-off between transmission traffic and required storage compare to I-frame and SP-frame.

Then, an unconstrained Lagrangian problem among the probability of switching views and storage capacity is defined. Redundant P-frames are added to frame structure instead of DSC-frame to solve the Lagrangian problem. Using redundant P-frame has lower transmission traffic than using DSC-frame. However, as the increase of the number of redundant P-frame, the storage cost will increase.

When the Lagrangian problem is resolved, the optimized frame structure is build. And then, server transmits the frames depending on feedback from client. [4] shows this method achieves better trade-off between transmission traffic and storage using I-, P-, and DSC-frame than only I-frame and I-, and P-frame.

2.5 UDMVT(User Dependent Multi-view Video Transmission)

SCC and MVC send all views and many frames are unnecessary. The bit-rate is high. UDMVT [5][6] is transmission technology that analyze the client's motion to prevent the transmission of redundant frames.

UDMVT focuses on the successive motion switching model in which client is just able to switch from current view to the neighboring views as shown by Fig. 2. In other word, if the multi-view video contains the views (1, 2 ... M), for any view j the client is just able to switch from j to the view j', where $max(1, j-1) \leq j' \leq min(j+1, M)$. In the successive motion model, which frames should be displayed when the client starts to switch to next view are decided by both the frame rate (frame/sec) of the multi-view video and the switching speed (view/sec) of the client. 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 client.

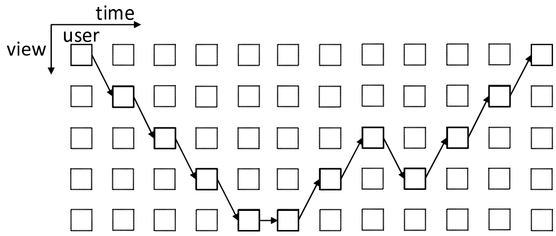

Figure 2: Successive motion model.

Therefore, if the client periodically fed back the three-tuples $N(p, f, s)$, server can predict an area in which frames may be displayed in next period of time as shown by Fig. 3. p is the initial position $F_{i0,j0}$ which is the frame of view j_0 at time i_0. f is the frame rate while s is the switching speed. Although this three-tuples cannot predict all the frames exactly in display path, it is able to predict a triangle area in which the frames are possible to be displayed in 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 bit-rate for the transmission of multi-view video.

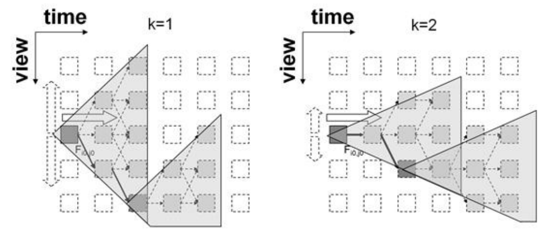

Figure 3: Feedback scheme for UDMVT.

UDMVT transmits only the necessary frames for the client according to the periodic feedback from client. As the increase of the number of views, UDMVT reduces more bit-rate than SCC and MVC. UDMVT reduce the more traffic of multi-view video with increasing number of views. The evaluation results show that up to 73.9% bit-rate is reduced when compared to SSC when the number of view is 8.

3 Problems of Multi-view Video in Wireless Network

Multi-view video streaming in wireless networks is attractive because of the convenience and mobility of wireless client terminals. With increasing number of clients and views, multi-view video streaming is affected by the bandwidth of the network, which brings a technical challenge to wireless streaming of multi-view video. Furthermore, due to the varying communication condition of wireless networks, if we do not deal about that variation the network latency and packet loss will increase and the frame error propagation may occur by the packet loss. The degradation can affect image quality and interactivity. These two factors are critical for providing video viewers with high quality videos.

In addition, the loads of server, networks and client terminals is also important factor. Especially, the loads of networks and client terminals have to be discussed because of the limitation of bandwidth in wireless networks, battery life and processing capabilities of client terminals. However, if the video is compressed too much to reduce the loads of networks and client terminals, the load of server is overwhelmed because of encoding process.

The following subsections discuss the detail of the problem of streaming multi-view video through the wireless networks.

3.1 Image quality and Interactivity

3.1.1 Network Latency

Network latency in multi-view video streaming is one of the key issues. When the network latency is large, interactivity of multi-view video decreases because if clients switch to other views, it will take long time for client terminals to feed back for the new views. Especially, for live streaming of multi-view videos, all frames should be transmitted to terminals within a limited latency. Typically, acceptable latency of live streaming via Internet is up to 300ms [4].

[12] discusses how much impact the network latency poses on the QoE (Quality of Experience) of server oriented and client oriented streamings. Server oriented streaming such as FTV means that server generates the perspective image from multi-view videos and transmits it to client. Client oriented streaming means server transmits the multiple videos and several information such as depth map [13][14] to clients. The client terminal generates the perspective image from the videos and the information. It evaluates the image quality and interactivity of server oriented and client oriented streamings with the average of MOS (Mean Opinion Score). They show that according to the increase of the network latency, the interactivity of server oriented streaming is drastically degraded. In client oriented streaming, the image quality is degraded significantly as the increasing of network latency. Although there is the difference of degradation between the two streamings, the network latency affects image quality and interactivity.

3.1.2 Packet Loss

H.264 video coding schemes use motion-compensated prediction (MCP) to achieve high compression efficiency on the expense of the error resilience. However, H.264 stream poses a severe frame error propagation problem. In wireless networks, the packet loss increases due to the errors in the physical layer. When a multi-view video is broadcasts to multiple clients, it incurs more packet loss because of interference. If a packet containing anchor frames such as I-frames and P-frames is lost, the errors would propagate to all dependent frames, leading to perceptible visual artifacts. [15] proposed an analysis-by-synthesis technique to estimate the perceptual value of each frame type. In [15], anchor frames are suggested to have better perceptual value than dependent frames.

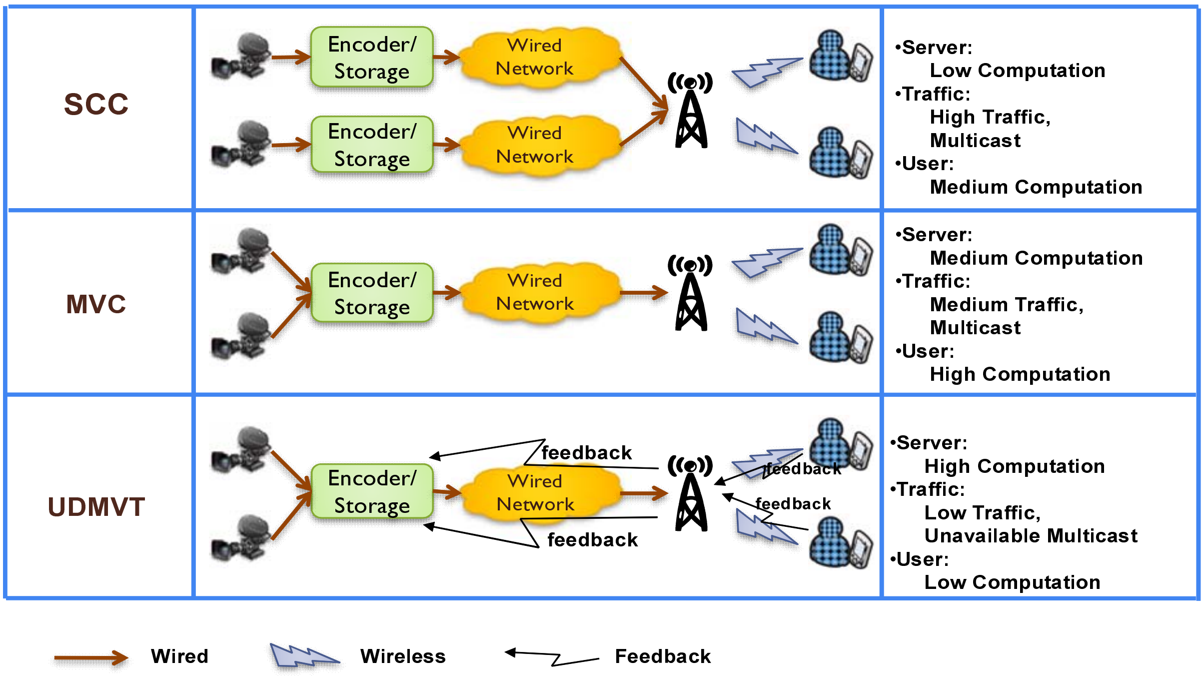

Figure 4: Characteristic of SCC, MVC and UDMVT.

3.2 The Loads of Server, Networks and Client Terminals

In wireless networks, there are the limitations of bandwidth and battery life and processing capabilities of a client terminal. Reducing the load of the networks and a client terminal is crucial for wireless networks. Also, the load of server leads to the load of networks and the client terminal. If the video is compressed too much at the server, although the load of networks (it means transmission bit-rate) decreases, much power will be required for decoding at the client terminal. When the server is not compressed the multi-view video, the much transmission bit-rate is required. Furthermore, a client terminal needs much power for receiving frames instead of decoding them. To reduce the processing at a client terminal, [16] proposes a framework of 3DTV streaming for mobile networks in which mobile terminal has low processing capabilities. The paper claims that the server should perform processing as much as possible and the client does little.

We summarize the discussion in Fig. 4 when using SCC, MVC and UDMVT in wireless networks. We assume that the access point and the encoder are connected with wired networks, ant that each client and the access point is connected with wireless networks.

Since SCC encodes each view independently, the server processing of encoding is little. Client processing of decoding is also little. However, since SCC encodes the frames without removing the inter-view redundancy, the transmission bit-rate is very large. MVC has a lower transmission bit-rate than SCC because it reduces the inter-view redundancy. However, the traffic is still high by transmitting all views. Besides, in order to display the multi-view video correctly, the frames which are displayed and the frames which depend on them must be received and decoded at first. Therefore, MVC consumes much power for decoding at the client terminal. UDMVT predicts a triangle area in which the frames are possible to be displayed in the next period of time by a client feedback. Thus, UDMVT increases the amount of processing of the server. It has a lower transmission bit-rate than SCC and MVC because it only encodes and transmits the potential frames. UDMVT needs less processing of decoding due to less number of frames to be decoded. Therefore, it reduces the power consumption and processing of a client terminal.

As mentioned above, SCC and MVC can reduce the load of the multi-view video server instead of increasing the load of networks and the client terminal. On the contrary, UDMVT reduces the load of networks and the client terminal while it causes larger load of server than SCC and MVC. In wireless networks, UDMVT is more suitable than SCC and MVC.

However, as the number of clients is increasing, server will be overwhelmed because there are the massive response and feedback from the clients. To prevent the overhead of the server, the server can use SCC and MVC to encode the images because these techniques encode frames of all views at once for the all clients on live streaming. Furthermore, the overall transmission bit-rate of SCC and MVC on live streaming does not increase because these techniques multicast the frames of all views to multiple clients. Therefore, we need to use the SCC, MVC and UDMVT together for multiple clients.

4 Two-step Encoding Frameworks

In this section, we present a framework for multi-view video streaming in wireless networks taking into account the benefits of SCC, MVC and UDMVT. Fig. 5 shows one of the frameworks using two-step encoding for reducing the latency,

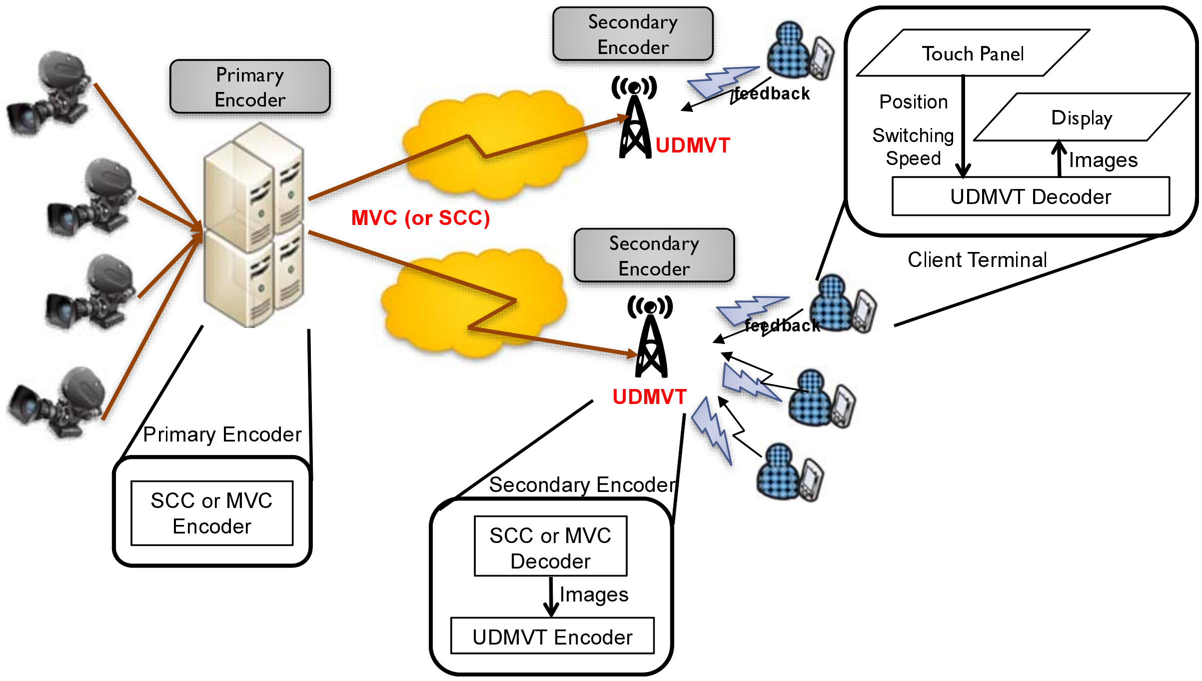

Figure 5: Proposed framework of multi-view video streaming in wireless networks

processing cost and packet loss in wireless networks. Although two-step encoding is similar technique to I-TCP[17] at transport layer, it is in application layer.

In two-step encoding, video frames taken by several cameras are encoded by the primary encoder using SCC or MVC. The primary encoder transmits the encoded frames to the secondary encoder via wired networks. In the secondary encoder, when the encoded frames are received, these frames are decoded to the original images using SCC or MVC decoder. Then, these original frames in a triangle area predicted by client's periodical feedback using UDMVT are encoded. Then, the secondary encoder transmits the frames to each client through wireless networks. After decoding the received frames, clients can watch the multi-view video. Further, each client feedback to the secondary encoder periodically.

4.1 Primary Encoder

Primary encoder encodes the images of multi-view video using SCC or MVC. When the primary encoder has a lower processing capability and higher bandwidth, it encodes the images with SCC. Otherwise, it encodes the images with MVC which requires more processing. Although the transmission bit-rate of multi-view video is high, the frames of all views can multicast to several networks at the sometime it they are encoded by SCC or MVC. This is especially useful for live streaming for multiple clients.

If the frames are encoded by UDMVT in the primary encoder, the transmission bit-rate can be reduced. However, UDMVT predicts and transmits triangle areas according to each client's feedback. With the increasing of clients, the processing of server will be very high. Furthermore, with increasing the latency through the several networks, UDMVT will not be able to respond client's feedback immediately. As a result, interactivity of multi-view video is lost. At the same time, UDMVT cannot multicast a triangle to multiple clients as SCC and MVC.

After encoding by SCC or MVC, primary encoder multicast the frames of all views to the secondary encoder on each networks that clients are watching the multi-view video.

4.2 Secondary Encoder

The secondary encoder is located at a base station or between the wired network and a base station of each wireless network. This encoder transmits frames to each client. When this encoder receives the encoded frames from primary encoder, these frames are decoded to the images using SCC or MVC decoder. After decoding, the secondary encoder re-encodes the images using UDMVT according to the feedback of each client. This is because UDMVT has lower power consumption for receiving and decoding frames at client terminal and also less transmission bit-rate than SCC and MVC. After re-encoding, the secondary encoder unicasts the potential frames to each client.

However, this process brings increasing of encoder's processing because the secondary encoder has to decode and re-encode the UDMVT frames. One of the ways to improve the efficiency is that SCC or MVC frame can be transformed to UDMVT frame at the secondary encoder directly. However, challenge is still remaining. For example, we need to discuss how to update the correlation between the frames. It means that how to remove the correlation from the encoded frames. Similarly, it means that how to add the new correlation to the encoded frames. Furthermore, this way needs to encode the different types of frame from received frames without decoding. For example, when the receiving frame is I-frame, it isn't so difficult to encode it to P-frame or B-frame, directly. However, when the receiving frame is a P-frame or B-frame, it is difficult to encode it to I-frame because redundant information is removed.

When the multiple clients are watching the same video in the same networks, the frames encoded for one client cannot be used by other clients in UDMVT. Many duplicate frames are encoded and transmitted. Therefore, as the increasing of clients in number, the overall transmission bit-rate and the secondary encoder's processing become higher. In SCC and MVC, the overall transmission bit-rate does not increase on live streaming because these techniques multicast the frames of all views to multiple clients. However, SCC and MVC have higher power consumption for each client because the transmission bit-rate and decoding processing for each client are still high. By the presented framework, the primary encoder using the SCC and MVC can reduce the increasing of overall transmission bit-rate with the increasing clients. The secondary encoders can release the high processing for server and reduce the power consumption of client terminal.

4.3 Client Terminals

Considering the battery life and processing capabilities of client's wireless terminals, the client structure needs to be simple. Smart phone is a good candidate such as Android and iPhone because it has a long battery life and can purchase low prices. Also, it is equipped with the touch panel standardly, it is suitable for the successive motion model which is requested by UDMVT.

When the client receives the encoded frames on interface software, these frames are decoded by UDMVT decoder. Therefore, each client can watch the streaming of multi-view video on wireless terminal. Furthermore, a client can switch to the different views using touch panel. Then, client's terminal feed back to the secondary encoder periodically.

5 Evaluations and Discussion

In this evaluation, we focus on the communication between secondary encoder and clients to evaluate the network latency and packet loss of multi-view video streaming through wireless networks. Essentially, we need to evaluate the latency of primary server to client terminals, the load of server and client terminals, image quality of video, overhead of re-encoding at the secondary server and so on. However, this paper only evaluates SCC, MVC and UDMVT for secondary encoder.

We use the multi-view video test sequence "ballroom" with 8 views and 320 × 480 resolution to evaluate the performance. This test sequence is provided by MERL [18]. The reference techniques are SCC, MVC and UDMVT. Encoders implemented by the modified open source project JMVC [19] is used to encode the multi-view video sequences. In each view, 250 frames were encoded with the frame rate 25 (fps). The length of the GOP was set as 8. Three values of the k were used: 5 to 15. The value of k means that how fast is

Table 1: Simulation parameter

Properties	Value
Distance between encoder and clients	100m
Data rate	11Mbps (802.11b)
Bandwidth	1Mbps
Packet size	1500bytes
Encoder's Buffer size	100Mbytes
The number of frames	3000frames

client switching views. When the value of k is 5, the client switches views quickly. As the value of k increases, it means the switching speed of client becomes slow. The length of feedback period in UDMVT were set as 60 (frames) and 120 (frames).

Then, we evaluate the network latency and packet loss by Qualnet network simulater [20] and the data of encoded frames. We assume the client switches among 25 views during the playback of 3000 frames. The simulation parameters of Qualnet are shown in Table 1.

5.1 Traffic of Multi-view Video

Fig. 6 and Fig. 7 show the transmission bit-rate of multi-view video streaming using each technique, when the feedback period of UDMVT is 60 and 120, respectively. Then, each client switches views randomly in 3000 frames. This simulation is performed 10 times, and then obtains the average of values. In both of these figures, SCC and MVC always have the same transmission bit-rate no matter how the client switches because they transmit all frames to client. The average of transmission bit-rate of SCC is 15462.2 (kbps) and

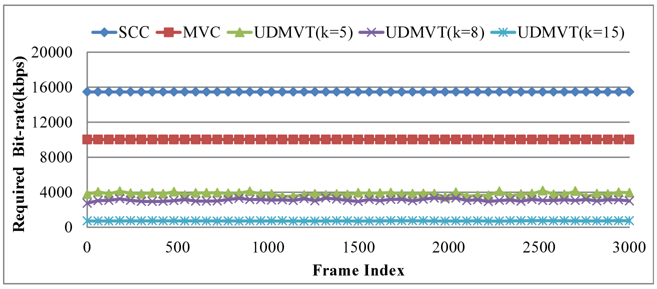

Figure 6: Transmission bit-rate of multi-view video when feedback period is 60(frames).

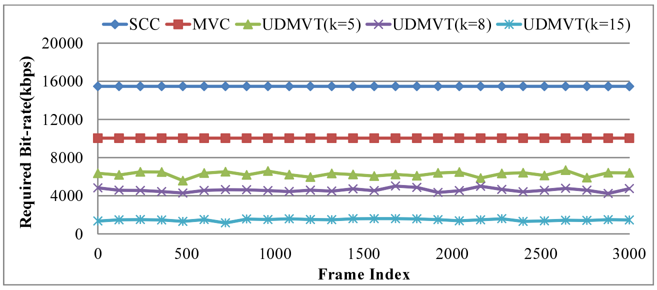

Figure 7: Transmission bit-rate of multi-view video when feedback period is 120(frames).

MVC is 10032.5 (kbps). On the other hand, when the client switches randomly, UDMVT predicts the potential frames by client's feedback and only transmit potential frames to client. Therefore, the transmission bit-rate of UDMVT varies depending on the switching speed. In Fig. 6, the average of transmission bit-rate of UDMVT (k=5) is 3837.9 (kbps), UDMVT (k=8) is 3103.3 (kbps) and UDMVT (k=15) is 732.3 (kbps) when the feedback period is 60. When the value of k is 5, which means client switches fast, the increase of the area of PFs will increase the transmission bit-rate. Then, with value of k increases, the frames need to be encoded and transmitted decrease because client switch views slowly. Thus, the transmission bit-rates of UDMVT are lower than MVC when the value of k is more than 5.

In Fig. 7, the average of transmission bit-rate of UDMVT (k=5) is 6262.5 (kbps), UDMVT (k=8) is 4599.0 (kbps) and UDMVT (k=15) is 1469.1 (kbps) when the feedback period is 120. If feedback period is increased, the frames which need to transmit to client are increased. Therefore, the transmission bit-rate of UDMVT is higher than the value when feedback period is 60.

5.2 Network Delay and Packet Loss of Multi-view Video Streaming

Fig. 8, 9, 10 and 11 show the average of network latency during the period the secondary encoder transmits frames to single client through wireless networks. SCC and MVC have large latency close to 1.3 seconds because they have a large transmission bit-rate to transmit on wireless networks. The average of network latency is 1.31 (sec) in SCC, 1.30 (sec) in MVC. It means that it is difficult to stream the multi-view video over wireless networks using these techniques. Fig. 8 also shows the average of latency of UDMVT is smaller than SCC and MVC because it has a lower traffic to transmit through the wireless networks when the feedback period is 60. Fig. 9 shows the detailed value of UDMVT. The average of latency of UDMVT (k=5) is 27.1 (ms), UDMVT (k=8) is 21.9 (ms) and UDMVT (k=15) is 5.2 (ms). This means that UDMVT does not affect the image quality and interactivity of multi-view video streaming through wireless networks in [12].

However, Fig.10 shows the average of latency of UDMVT (k=5) increases dramatically about 1.29 (sec) because its traffic is difficult to transmit the frames to client in wireless networks when the feedback period is 120. When the value of k increases, the average of latency is decreased because the required transmission bit-rate of UDMVT is reduced. Fig.11 shows the average of latency of UDMVT (k=8) is 32.9 (ms) and UDMVT (k=15) is 10.3 (ms). Therefore, even if the feedback period is increased, UDMVT can provide the streaming of multi-view video over wireless networks with high image quality and interactivity, when the client switch views slowly.

Fig. 12 and Fig. 13 represent the rate of packet loss when the secondary encoder transmits frames to single client through wireless networks. Even if the feedback period changes, the average of packet loss of SCC and MVC remains high about 66.8% in SCC and 48.8% in MVC. When the feedback period is 60, there is almost no packets loss in UDMVT as shown Fig. 12. However, when the feedback period increases, the packet loss of UDMVT (k=5) about 17.8% because its traffic is increased and afflicts the bandwidth of wireless networks as shown Fig. 13. Thus, if we assume an environment in which

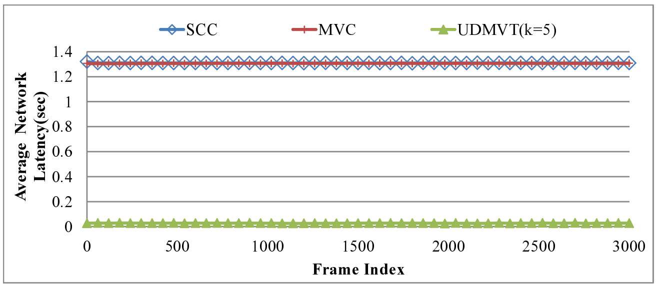

Figure 8: Network latency of multi-view video streaming on wireless network when feedback period is 60(frames).

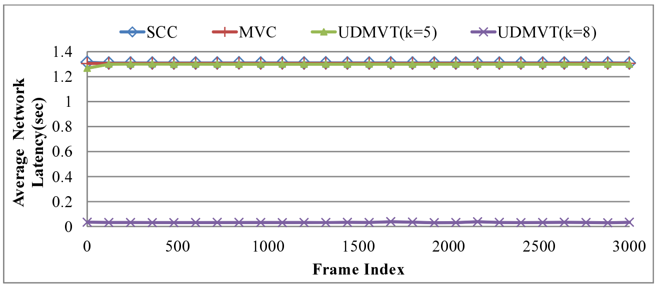

Figure 10: Network latency of multi-view video streaming on wireless networks when feedback period is 120(frames).

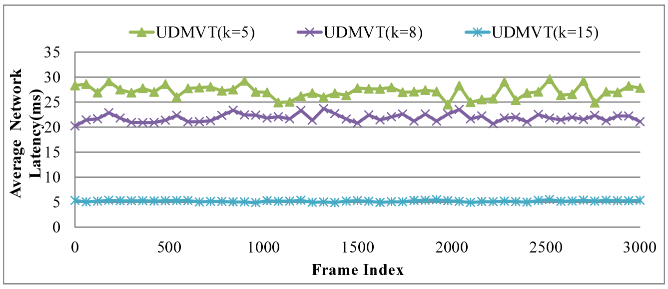

Figure 9: The detail of Network latency with UDMVT when feedback period is 60(frames).

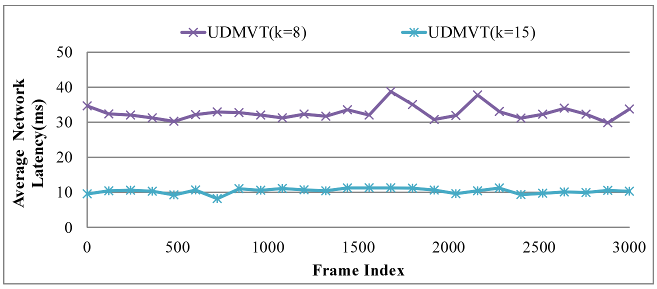

Figure 11: The detail of Network latency with UDMVT when feedback period is 120(frames).

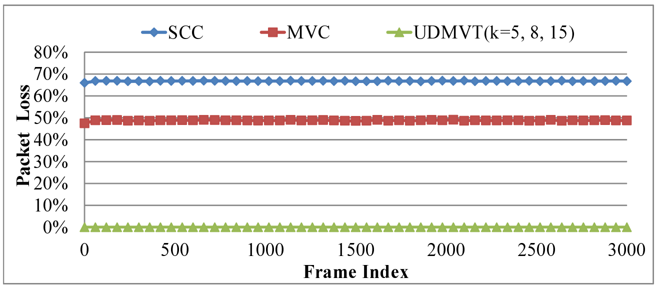

Figure 12: Packet loss of multi-view video streaming on wireless networks when feedback period is 60(frames).

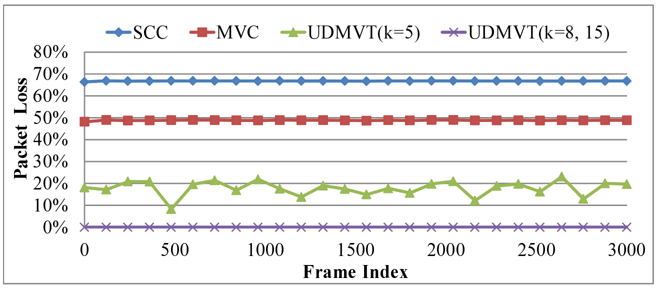

Figure 13: Packet loss of multi-view video streaming on wireless networks when feedback period is 120(frames).

client switch views at a high speed or long feedback period for multi-view video streaming over wireless networks, it is necessary to propose the new techniques based on UDMVT for wireless networks.

6 Conclusion

In this paper, we discuss the multi-view video streaming over wireless networks and present a framework to reduce the latency, processing cost and packet loss. The framework includes two-step encoding. On the first-step, the primary encoder encodes the multi-view video by SCC or MVC to reduce the overall transmission bit-rate with the increasing clients. On the second-step, secondary encoder decodes the received frames to images and then re-encodes the images using UDMVT. It can provide the high processing for the server and reduce the power consumption of client terminal. We evaluate SCC, MVC and UDMVT with respect to the transmission bit-rate of multi-view video, network latency and packet loss for all criteria the secondary encoder to end-user using network simulater Qualnet. Results show that UDMVT outperforms others.

In future works, we evaluate the overall performance of the two-step encoding. For example, there are the latency of the primary encoder to the client terminals, the loads of the server and the client terminals, image quality of videos and overhead of re-encoding at the secondary encoder. Furthermore, we compare the performance of two-step encoding framework with one-step encoding framework.

REFERENCES

[1] M. Tanimoto, "Overview of Free Viewpoint Television," Signal Processing: Image Communication, vol. 21, no. 6, pp. 454-461, July 2006.

[2] H. Kimata, M. Kitahara, K. Kamikura, Y. Yashima, T. Fujii, M. Tanimoto, "Low delay multi-view video coding for free-viewpoint video communication," IEICE Jpn. , J89-J (1), pp. 40-55, 2006.

[3] A. Vetro, P. Pandit, H. Kimata, A. Smolic and Y-K. Wang, "Joint Draft 8.0 on Multi-view Video Coding," Joint Video Team, Doc. JVT-AB204, July 2008.

[4] Xiaoyu, Xiu, Gene Cheung, Jie Liang, "Frame Structure Optimization for Interactive Multiview Video Streaming with Bounded Network Delay," IEEE International Conference on Image Processing, Nov. 2011.

[5] Ziyuan Pan, Yoshihisa Ikuta, Masaki Bandai, Takashi Watanabe "User Dependent Scheme for Multi-view Video Transmission", IEEE International Conference on Communications, June 2011.

[6] Ziyuan Pan, Yoshihisa Ikuta, Masaki Bandai, Takashi Watanabe, " A User Dependent System for Multi-view Video Transmission, " IEEE AINA2011, pp.732-739, Mar. 2011, Singapore.

[7] K. Mueller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Oelbaum, and T. Wiegand, "Multi-view video coding based on H.264/AVC using hierarchical B-frames," Picture Coding Symposium 2006, 2006.

[8] P. Merkle, K. Muller, A. Smolic, and T. Wiegand, "Statistical evaluation of spatiotemporal prediction for multi-view video coding," Proc. ICOB 2005, Berlin, Germany, pp. 27-28, Oct. 2005.

[9] A. Kaup and U. Fecker, "Analysis of multireference block matching for multi-view video coding," Proc. 7th Workshop Digital Broadcasting, Erlangen, Germany, pp. 33-39, Sep. 2006.

[10] E. Kurutepe, M.R. Civanlar, and A.M. Tekalp, "Client-driven selective streaming of multi-view video for interactive 3DTV," IEEE Trans. CSVT, Oct. 2007.

[11] Ngai-Man Cheung, Antonio Ortega, Gene Cheung, "Distributed Source Coding Techniques for Interactive Multiview Video Streaming," Picture Coding Symposium, 2009.

[12] Ayano Tatematsu, Bohai Liu, Norishige Fukushima and Yutaka Ishibashi, "Influence of Network Delay on Quality of Experience in Free-Viewpoint Video Transmission," ITEJ Trans., vol. 65, no. 12, pp. 1742-1749, Dec. 2011.

[13] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger," Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability ", Signal Processing: Image Communication. Special Issue on 3DTV, February 2007.

[14] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, " High-Quality Video View Interpolation Using a Layered Representation ", ACM SIGGRAPH and ACM Trans. on Graphics, Los Angeles, CA, USA, August 2004.

[15] F. De Vito, D. Quaglia and J. De Martin, "Model-based distortion estimation for perceptual classification of video packets," in Proc. IEEE Int. Conf. on Multimedia & Expo, vol. 1, pp. 141-144, August 2002.

[16] J. Kwon, M. Kim, and C. Choi, " Multiview Video Service Framework for 3D Mobile Devices, "in International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, pp. 1231-1234, 2008,

[17] Bakre, A., Badrinath, B.R., "I-TCP: indirect TCP for mobile hosts," IEEE International Conference on Distributed Computing Systems, Vancouver, Canada, pp. 136-143, May 1995.

[18] ISO/IEC JTC1/SC29/WG11, "Multiview Video Test Sequences from MERL", Doc.M12077, Busan.Korea , April 2005.

[19] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG. JMVC (Joint Multiview Video Coding) software, June 2008.

[20] Qualnet Network Simulator, http://www.scalable-networks.com.