A Preliminary Study on Multi-view Video Streaming over Underwater Acoustic Networks

Takuya Fujihashi*, Hai-Heng Ng†, Ziyuan Pan*, Shunsuke Saruwatari*, Hwee-Pink Tan‡ and Takashi Watanabe*

*Graduate School of Informatics, Shizuoka University, Japan

†Engineering Department, Institute for Infocomm Research, Singapore

‡Sense and Sense-abilities, Institute for Infocomm Research, Singapore

{fujihashi, pan}@aurum.cs.inf.shizuoka.ac.jp, {nghh, hptan}@i2r.a-star.edu.sg, {saru, watanabe}@inf.shizuoka.ac.jp

***Abstract*—Several techniques to realize multi-view video streaming have been recently proposed for terrestrial wireless networks. Multi-view video allows users to watch object of interest from different angle and can be applied to many applications. Nonetheless, the existing techniques do not cater for high-latency networks, such as underwater acoustic networks. Due to the slow speed of sound in water, the response delay of multi-view video becomes longer, which in turn degrades users' experience. This paper proposes a new approach, called Feedback-based Multi-view video Streaming for Underwater Acoustic Networks (FMS-UAN) so as to reduce the response delay of multi-view video streaming. FMS-UAN has two unique features. First, a sender can transmit its predicted frames before receiving feedback frame from a receiver node. We exploit long propagation delay by allowing both sender and receiver to simultaneously transmit their respective frames to each other in a collision-free manner. Second, even if the prediction was wrong, the sender still can encode the correct frame area with transmitted frames to reduce the retransmission bit-rate.**

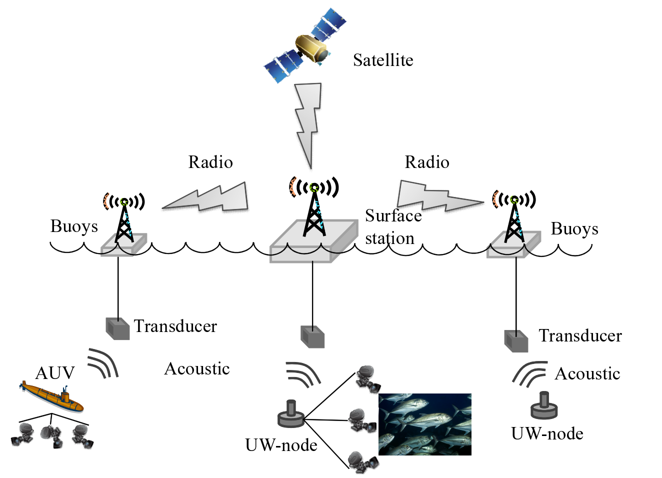

Fig. 1. Architecture of multi-view video streaming over underwater acoustic networks

I. Introduction

The prevalence of low-cost camera technology has spurred the development of multi-view video streaming in terrestrial networks, which offers users an opportunity to switch their viewpoints freely [1], [2]. We envision that this technique can also be applied in a wide spectrum of underwater applications, such as ocean observatory and underwater vehicle monitoring as shown in Fig. 1. As explained in [3], in video streaming applications, the response delay (i.e., the delay from sending feedback to receiving all video frames) is highly critical to achieve better user experience. As the response delay increases, user experience the intermittent delay of switching views and stuttering.

From the literature, there are two common methods for underwater single-view video streaming to decrease the response delay, namely: (i) tethered transmission [4] and (ii) acoustic wave communication [5]. Although the former ensures good quality video transmissions, it incurs higher deployment cost and is often less flexible than the latter. Compared to single-view video transmission, multi-view video has significantly higher traffic volume. Furthermore, previous multi-view video research only handles wired network environment which has a very high data rate and negligible propagation delay. Thus, it is even more challenging to realize *multi-view* streaming in high-latency and bandwidth-limited underwater acoustic networks.

To decrease the response delay of multi-view video streaming via underwater acoustic networks, we propose a novel Feedback-based Multi-view video Streaming for Underwater Acoustic Networks (FMS-UAN) approach. In FMS-UAN, an intended receiver periodically sends feedback to a sender, which conveys information such as current view frame and switching speed. Based on the feedback information, the sender only needs to encode and transmit specific video frames that are required by the receiver, which greatly reduces the bit-rate of video transmission. More specifically, FMS-UAN has two unique features: Predictive Video Transmission and Reusable Encoding.

First, a sender can periodically predict an initial view position and switching speed of the receiver, before transmitting several video frames that are likely to be playback by the receiver. Note that, to better exploit long propagation delay in underwater networks, we allow both sender and receiver nodes to simultaneously transmit their respective video frames and feedback packets in a collision-free manner. Second, FMS-UAN uses a new multi-view frame prediction structure, which helps to reduce the bit-rate of video frame retransmissions if the predicted frame area is less accurate. In this structure, the correct frame area is encoded using transmitted frames. To our best knowledge, we are the first to study multi-view video streaming in underwater acoustic networks.

The rest of this paper is organized as follows: In Section Ⅱ, we review existing works of underwater video streaming and multi-view video streaming. Furthermore, we also explain challenges and problems of multi-view video streaming over underwater acoustic networks. Based on Section Ⅱ, we propose FMS-UAN approach so as to reduce the response delay in Section Ⅲ. Finally, Section Ⅳ gives the conclusion.

II. Related Work

A. Application of Underwater Video Streaming

In conventional research, useful information, such as water temperature and quality, is transmitted from senders to receiver in underwater acoustic network. Underwater video streaming can be applied in several applications, such as environmental monitoring, disaster prevention and assisted navigation [6], [7].

Our research assumes the sender has video cameras and transmits single video or multiple videos to the receiver through underwater acoustic network. By transmitting the video, this topology can be applied to a variety of underwater applications. Hereinafter, we describe applications of single-view video and multi-view video streaming for underwater.

1) Single-view Video: At first, we describe the case of single-view video streaming. In this case, the sender located in underwater has one or multiple cameras and then transmits only one video to the receiver. User can watch the current situation on display because the video provides perceivable information to the user. Furthermore, the user can notice even minor changes. By taking this advantage, many applications such as fish observation and disaster prevention are enable [8].

2) Multi-view Video: We then describe the case of multi-view video streaming for underwater. The sender has multiple cameras and transmits multiple videos to the receiver in this case. User can watch objects from every angle by receiving multiple videos. So far, there is no research to deal with multi-view video streaming for underwater acoustic network. However, we think our research can be applied ocean observatory, fish farming and underwater vehicle monitoring.

B. Underwater Video Streaming

Previous research of underwater video streaming can be classified two types: tethered transmission and acoustic wave communication. Shiau [4] achieves a high-resolution video streaming via optical fiber under the water. In this paper, the sender located under the water transmits the video with the resolution of 1280 $\times$ 1080 pixels to the receiver located on the sea. Although optical fiber ensures good quality video transmissions, it incurs higher deployment cost and is often less flexible than the video streaming over acoustic network.

The previous studies in [5], [9], [10], [11] are related paper for video streaming over underwater acoustic network. In fact, acoustic network has a much lower data rate than radio network (i.e. 20 kbps in acoustic network and 2 Mbps in radio network of IEEE 802.11b). This data rate is difficult to stream the video because the video has a high bit-rate and the video cannot be decoded if the data rate is below the encoded video bit-rate. In [5], [9], [10], the authors use higher data rate underwater modem to perform video streaming. [5] achieves 90 kbps in 115 kHz acoustic band over 200 meters vertical link under a variety of channel conditions using OFDM modulation in acoustic network. Furthermore, [9] achieves 150 kbps in 75 kHz acoustic band over 8 meters vertical link using 64-QAM modulation.

Pompili and Akyildiz [11] proposes a cross-layer protocol for underwater multimedia applications. This protocol optimizes the resource allocation for routing, MAC, and PHY layer according to the different types of multimedia applications: delay-tolerant or delay-sensitive and loss-tolerant or loss-sensitive. Note that these researches handle only single-view video streaming.

In general, all the video frames from multiple video cameras are transmitted to the receiver in multi-view video streaming [14], [15]. However, the bit-rate is very high: about 5 Mbps for 704 $\times$ 480, 30 fps, and 8 camera sequences [12]. This bit-rate hardly decodes the video at the receiver. To reduce the bit-rate, the receiver transmits feedback information to the sender periodically [17], [18], [19]. To exploit the feedback information such as initial position and switching speed, the sender only replies limited video frames to the receiver.

However, if the data rate of acoustic network becomes higher and the bit-rate of multi-view video becomes lower, the response delay, which represents the delay from sending feedback to receiving all video frames, becomes higher in underwater acoustic network because of the long propagation delay. As the response delay increases, the user experiences the intermittent delay of switching views and stuttering.

C. Multi-view Video Streaming

We first describe how to visualize video frames in The Moving Picture Experts Group (MPEG) [13]. 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. 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. I-frame has a high bit-rate while P- and B-frame have a low bit-rate.

Previous research of multi-view video based on MPEG can be classified two types: all video sequences transmission [14], [15] and user dependent streaming [17], [18], [19].

1) All Video Sequences Transmission: In this case, the sender transmits all video sequences to the sender to avoid the switching delay. Typical scheme includes SimulCast Coding

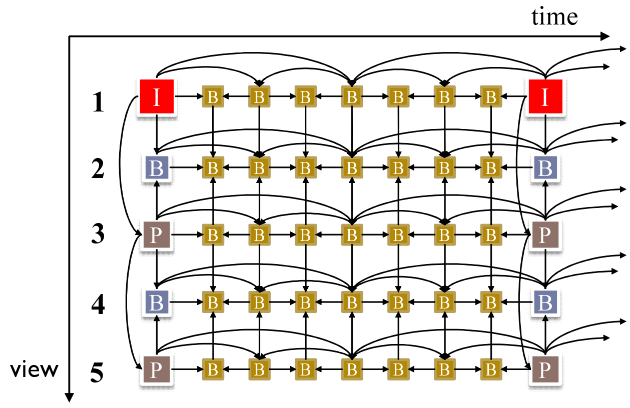

Fig. 2. Prediction structure of MVC

(SCC) and Multi-view Video Coding (MVC). The straight-forward solution for multi-view video encoding is SCC [14] in which all video sequences are encoded independently using H.264/MPEG-4 AVC compression technology. However, SCC encoded video still contains a large amount of inter-view redundant information.

A large amount of inter-view redundant information is still contained in SCC encoded video. In order to remove the correlation between views, MVC was issued as an amendment to H.264/MPEG-4 AVC [15]. MVC combines the temporal prediction and inter-view prediction as shown in Fig. 2. Statistical evaluations [20], [21] have shown significant compression gains. However, even with the MVC, bit-rate for multi-view video is still high.

2) User Dependent Streaming: In this type, the sender transmits only the limited frames requested by the receiver. Some schemes send the request to the sender on each frame, other schemes send the request the sender on several frames periodically.

In order to reduce the transmission bit-rate, Client-Driven Selective Streaming (CDSS) protocol, which combines MVC and scalable video coding (SVC), was proposed [16]. In this protocol, the view needed by the user is decided by a predictor. The server encodes the view into two quality levels: the base layer and the enhancement layer. The base layer encodes all views into a lower bit-rate by MVC, and the enhancement layer encodes the views which are selected by user to allow random access and improve the quality of the base layer. However, the performance of this system depends on the Kalman filter-based predictor. If there are no prediction errors, the high-quality streams are displayed. If the prediction is incorrect, only the base layer (low-quality) is displayed and it brings the bad client experience.

In [17], an algorithm of building optimized frame structure in order to reduce transmission traffic on Interactive Multi-view Video Streaming (IMVS) 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. This paper shows that this method achieves better trade-off between transmission bit-rate and storage.

User Dependent Multi-view Video Transmission (UDMVT) [18], [19] is an interactive live multi-view video streaming scheme that transmits only the necessary frames for the user according to the periodic feedback. In order to predict the user view-switching, UDMVT focuses on the successive motion switching model. In the successive motion model, the frames can be categorized as either potential frames (PFs) or redundant frames (RFs), depending on the value of k. Here, k is the floor of the frame rate divided by the current switching speed, i.e., $k = \lfloor f/s \rfloor$, where f denotes the frame rate and s denotes the current switching speed of the user. The basic concept behind UDMVT is that only the PFs are encoded and transmitted, whereas the RFs are ignored. Therefore, UDMVT achieves lower bit-rate than MVC.

To extend the multi-view video streaming for multiple users, [22] is proposed based on UDMVT. When the sender tries to transmit the multi-view video to multiple users, many frames are overlapped among multiple users. This problem increases the redundant transmission of video frames and the overall bit-rate. User dependent Multi-view video Streaming for Multi-users (UMSM) reduces the overall bit-rate by multicasting the overlapping frames among multiple users.

Layered User dependent Multi-view video Streaming (LUMS) [23] is an extension of UDMVT. If the prediction of the user's switching is incorrect, then the user dependent streaming will have a large retransmission bit-rate. LUMS optimizes the prediction structure for retransmission of video frames. This structure reduces the retransmission bit-rate.

However, these schemes assume wired network which has a very high data rate and negligible propagation delay (i.e. data rate is 1 Gbps and physical speed depends on the speed of light). So far, there is no scheme to handle a long delay in multi-view video streaming.

III. Our Proposed FMS-UAN Scheme

In underwater acoustic network, the response delay becomes higher because of long propagation delay and thus user experience intermittent delay when switching views and stuttering. To decrease the response delay, we propose a novel approach called FMS-UAN. Fig. 3 shows the timing diagram of proposed FMS-UAN. We assume an intended receiver node is located on the sea surface, while the initiating sender is located at the seabed. Furthermore, the feedback interval (frames) is fixed value determined by the sender.

When the user begins to watch the multi-view video, the receiver sends the feedback packet to the sender. The feedback includes initial view and view-switching speed of the user. The sender encodes the video frames according to the feedback with encoding delay (t^E) when the sender receives the feedback packet. Then, it transmits the encoded video frames and notifies the feedback interval to the receiver. When the receiver receives the frames, the receiver decodes it with decoding delay (t^D) and then the user watches the multi-view video on display.

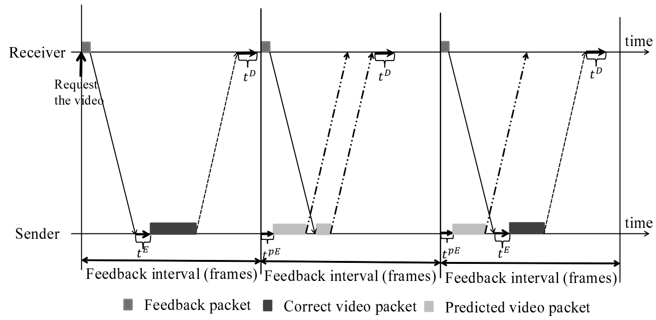

Fig. 3. Timing diagram of FMS-UAN

After a while, FMS-UAN starts the next feedback in order to playback the multi-view video continuously. To decrease the response delay and the user does not experience the intermittent delay of switching views and stuttering, FMS-UAN's design is based on two key components: 1) Predictive Video Transmission and 2) Reusable Encoding.

1) *Predictive Video Transmission*: At the beginning of the next feedback, the receiver sends the feedback information to the sender. At the same time, the sender predicts the initial frame area with predicted encoding delay (t^{pE}). Then, the sender encodes and transmits the predicted frame area to the receiver until receiving the feedback. We call this method Predictive Video Transmission. Note that a collision between feedback packet and video packet happens at the sender in radio network. It means the sender cannot receive the feedback and encode the video frames using it. Moreover, the receiver cannot receive complete video packets and watch the video on display. Therefore, the sender and receiver need to re-transmit the packets. As a result, the response delay becomes longer.

Similar to the bidirectional-concurrent transmission approach proposed in [24], we also allow both sender and receiver nodes to simultaneously exchange their respective predicted video frames and feedback packets at the starting of each feedback cycle (as shown in Fig. 3), so as to better exploit long propagation delay characteristic. Since the simultaneous packet exchange happens in a collision-free manner, this helps to further reduce the response delay. If the prediction is correct, the sender does not need to re-encode the video and only transmits the rest of frames after receiving the feedback. Hence, FMS-UAN can decrease the response delay drastically. In this case, achieving high prediction accuracy leads to a reduction of the average response delay. To predict the frame area with high prediction accuracy, FMS-UAN uses three types of solution depending on the switching models of user: Kalman filter, Bayesian estimation and Prediction by partial matching. Therefore, FMS-UAN achieves high prediction accuracy in accordance with the movement of the user. While our scheme assumes periodical transmission cycle, in which time synchronization between sender-receiver node pair is crucial, we can easily achieve good time synchronization, such as using the method proposed in [25].

2) *Reusable Encoding*: After receiving the feedback, the sender confirms that the predicted frame area was correct. If the prediction was correct, the sender continues to transmit the rest of frames to the receiver. However, the sender needs to encodes the video based on the feedback if the prediction was wrong. It incurs an extra overhead of encoding and increases the load of the sender. In FMS-UAN, the sender encodes a correct frame area according to the feedback and the frames which have been transmitted if the prediction includes wrong part. Then, the sender transmits the correct frames to the receiver. As a result, FMS-UAN reduces the retransmission bit-rate of multi-view video by exploiting the transmitted frames.

IV. Conclusion

This paper proposes a novel approach of FMS-UAN to reduce the response delay of multi-view video streaming for underwater acoustic networks. FMS-UAN's design is based on two key components: Predictive Video Transmission and Reusable Encoding. First, a sender can transmit its predicted frames before receiving feedback frame from a receiver node. We exploit long propagation delay by allowing both sender and receiver to simultaneously transmit their respective frames to each other in a collision-free manner. Second, even if the prediction was wrong, the sender still can encode the correct frame area with transmitted frames to reduce the retransmission bit-rate.

References

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

[2] H. Kimata, M. Kitahara, K. Kamikura, Y. Yashima, T. Fujii, M. Tanimoto, "Low-delay multiview video coding for free-viewpoint video communication," *Systems and Computers in Japan*, Vol. 38, no. 5, pp. 14-29, May 2007.

[3] A. Tatematsu, B. Liu, N. Fukushima, Y. Ishibashi, "Influence of Network Delay on Quality of Experience in Free-Viewpoint Video Transmission," *The journal of the Institute of Image Information and Television Engineers*, vol. 65, no. 12, pp. 1742-1749, Dec. 2011.

[4] Y. H. Shiau, S. I. Lin, F. P. Lin, C. C. Chen, "Real-Time Fish Obesrvation and Fish Category Database Construction," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 4, pp. 45-49, 2012.

[5] J. Ribas, D. Sura, M. Stojanovic, "Underwater Wireless Video Transmission for Supervisory Control and Inspection using Acoustic OFDM," *in Proc. IEEE OCEANS*, pp. 1-9, Sep. 2010.

[6] I. F. Akyildiz, D. Pompili, T. Melodia, "State-of-the-Art in Protocol Research for Underwater Acoustic Sensor Networks," *in Proc. ACM WUWNet*, pp. 7-16, Sep. 2006.

[7] M. Suzuki, T. Sasaki, T. Tsuchiya, "Digital Acoustic Image Transmission System for Deep-sea Research Submersible," *in Proc. IEEE OCEANS*, pp. 567-570, Oct. 1992.

[8] D. Pompili, I. F. Akyildiz, "A cross-layer communication solution for multimedia applications in underwater acoustic sensor networks," *in Proc. IEEE MASS*, pp. 275-284, Sep. 2008.

[9] C. Pelekanakis, M. Stojanovic, L. Freitag, "High Rate Acoustic Link for Underwater Video Transmission," *in Proc. IEEE OCEANS*, pp. 1091-1097, Sep. 2003.

[10] L. D. Vall, D. Sura, M. Stojanovic, "Towards Underwater Video Transmission," *in Proc. ACM WUWNet*, pp. 1-5, Sep. 2011.

[11] D. Pompili, I. F. Akyildiz, "A Multimedia Cross-Layer Protocol for Underwater Acoustic Sensor Networks," *IEEE Trans. on Wireless Communications*, vol. 9, no. 9, pp. 2924-2933, Sep. 2010.

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

[13] D. Marpe, T. Wiegand, G. J. Sullivan, "The H. 264/MPEG4 advanced video coding standard and its applications," *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134-143, Aug. 2006.

[14] Joint Video Team ISO/IEC JTC1/SC29/WG11, "Updated call for proposals on multi-view video coding," *MPEG2005/N7567*, Oct. 2005.
[15] K. Mueller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Oelbaum, T. Wiegand, "Multi-view video coding based on H.264/AVC using hierarchical B-frames," *in Proc. IEEE PCS*, pp. 385-390, Apr. 2006.
[16] E. Kurutepe, M. R. Civanlar, A. M. Tekalp, "Client-driven selective streaming of multi-view video for interactive 3DTV," *IEEE Trans. CSVT*, vol. 17, no. 11, pp. 1558-1565, Oct. 2007.
[17] X. Xiu, G. Cheung, J. Liang, "Frame Structure Optimization for Interactive Multiview Video Streaming with Bounded Network Delay," *in Proc. IEEE ICIP*, pp. 593-596, Sep. 2011.
[18] Z. Pan, Y. Ikuta, M. Bandai, T. Watanabe, "User Dependent Scheme for Multi-view Video Transmission", *in Proc. IEEE ICC*, pp. 1-5, Jun. 2011.
[19] ——————, "A User Dependent System for Multi-view Video Transmission," *in Proc. IEEE AINA*, pp.732-739, Mar. 2011.
[20] P. Merkle, K. Muller, A. Smolic, T. Wiegand, "Statistical evaluation of spatiotemporal prediction for multi-view video coding," *in Proc. ICOB* , pp. 27-28, Oct. 2005.
[21] A. Kaup, U. Fecker, "Analysis of multireference block matching for multi-view video coding," *in Proc. 7th Workshop Digital Broadcasting*, pp. 33-39, Sep. 2006.
[22] T. Fujihashi, Z. Pan, T. Watanabe, "Traffic Reduction for Multiple Users in Multi-view Video Streaming," *in Proc. IEEE ICME*, pp. 182-187, Jul. 2012.
[23] Z. Pan, M. Bandai, T. Watanabe, "Layered User Dependent Multi-view Video Streaming," *in Proc. IEEE PCS*, pp. 89-92, May 2012.
[24] H. H. Ng, W. S. Soh, M. Motani, "BiC-MAC: Bidirectional-Concurrent MAC Protocol with Packet Bursting for Underwater Acoustic Networks," *in Proc. MTS/IEEE OCEANS*, Sep. 2010.
[25] A. A. Syed, J. Heidemann, "Time Synchronization for High Latency Acoustic Networks," *in Proc. IEEE INFOCOM*, pp. 1-12, Apr. 2006.