ENHANCING THE SMOOTHNESS OF STREAMING VIDEO FOR MOBILE USERS OVER UNRELIABLE NETWORKS

Hussein Muzahim Aziz

Blekinge Institute of Technology
Licentiate Dissertation Series No. 2010:09

School of Computing

Enhancing the Smoothness of Streaming Video for Mobile Users over Unreliable Networks

Hussein Muzahim Aziz

Blekinge Institute of Technology Licentiate Dissertation Series
No 2010:09

Enhancing the Smoothness of Streaming Video for Mobile Users over Unreliable Networks

Hussein Muzahim Aziz

School of Computing
Blekinge Institute of Technology
SWEDEN

School of Computing
Publisher: Blekinge Institute of Technology
Printed by Printfabriken, Karlskrona, Sweden 2010
ISBN: 978-91-7295-190-7
Blekinge Institute of Technology Licentiate Dissertation Series
ISSN 1650-2140
urn:nbn:se:bth-00476

Abstract

Real time video streaming over wireless networks is an increasingly important and attractive service to the mobile users. Video streaming involves a large amount of data to be transmitted in real time, while wireless channel conditions may vary from time to time. It is hard to guarantee a reliable transmission over the wireless network, where the parameters specifying the transmissions are; bandwidth, packet loss, packet delays, and outage times. The quality of the video is affected negatively when network packets are lost, and the mobile users may notice some sudden stop during the video playing; the picture is momentarily frozen, followed by a jump from one scene to a totally different one.

The main objective of this thesis is to provide a smooth video playback in the mobile device over unreliable networks with a satisfactory video quality. Three different techniques are proposed to achieve this goal. The first technique will stream duplicate gray scale frames over multichannels, if there is lost frames in one channel it can be recovered from another channel. In the second technique, each video frame will be split into sub-frames. The splitted sub-frames will be streamed over multichannels. If there is a missing sub-frame during the transmission a reconstruction mechanism will be applied in the mobile device to recreate the missing sub-frames. In the third technique, we propose a Time Interleaving Robust Streaming (TIRS) technique to stream the video frames in different order. The benefit of that is to avoid the losses of a sequence of neighbouring frames. A missing frame from the streaming video will be reconstructed based on the surrounding frames.

The Mean Opinion Score (MOS) metric is used to evaluate the video quality. The experienced quality of a video is subject to the personal opinion, which is the only goal to satisfy the average human watching the contents of the video.

Acknowledgements

I would like to express my gratitude to Professor Lars Lundberg for his supervision, encouragement, and suggestions throughout the development of this thesis.

Special thanks to Professor Håkan Grahn for his guidance, advice, support and invaluable discussions that encouraged me in my research. I am also very grateful to Dr. Markus Fiedler for his guidance and valuable critical feedback at every stage of this study.

I would like to thank all my colleagues at the School of Computing for a friendly and enjoyable working environment. I would like to thank the administration staff for their kind assistance, and the library staff for helping me to set up the environment study.

Finally, I would like to thank my family and friends for their encouragement and support. Without them I would never come up to this stage.

Karlskrona, November 2010

Hussein Aziz

This work is dedicated to Yosser

List of Publications

The thesis is based on the following publications:

1. Hussein Muzahim Aziz, Lars Lundberg. Graceful Degradation of Mobile Video Quality over Wireless Network. IADIS International Conference Informatics, Algarve, Portugal, June 2009, pp. 175-180. As chapter 2.

2. Hussein Muzahim Aziz, Håkan Grahn, Lars Lundberg. Eliminating the Freezing Frames for the Mobile User over Unreliable Wireless Networks. Sixth International Conference on Mobile Technology, Applications and Systems. ACM Mobility Conference, Nice, France, September 2009, pp. 57:1 - 57:4. As chapter 3.

3. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Streaming Video as Space – Divided Sub-Frames over Wireless Networks. Third Joint IFIP Wireless and Mobile Networking Conference, WMNC'10, Budapest, Hungary, October 2010. As chapter 4.

4. Hussein Muzahim Aziz, Håkan Grahn, Lars Lundberg. Sub-Frame Crossing for Streaming Video over Wireless Network. The Seventh International Conference on Wireless On-demand Network Systems and Services, WONS’10, Kranjska Gora, Slovenia, February 2010, pp: 53-56. As chapter 5.

5. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Eliminating the Effects of Freezing Frames on User Perceptive by Using a Time Interleaving Technique. Submitted for journal publication. As chapter 6.

Table of Contents

Chapter Three

Chapter Four

Chapter Five

Chapter Six

CHAPTER ONE

1.1 Motivation

Real time video streaming over wireless networks often suffers from unacceptable delay, jitter and packet loss. Among these, packet loss often has the largest impact on the quality of the video. Our research goal is to provide a satisfactory video quality for real time video streaming over unreliable networks.

To identify the issues in a streaming video system we need to study the components of the streaming video system. The failure could be in the streaming server, in the wireless network, or in the mobile devices. The main focus of this thesis is to study the effects of frame losses during the video transmission in real time.

Streaming video in real time requires special techniques that can overcome the losses of packets in the unreliable networks. In this thesis, we propose several techniques to overcome the network limitations and provide a smooth video quality to be played on the mobile device even when there are high frame loss rates.

1.2 Background

1.2.1 Video Coding

Video coding using the H.264 / MPEG-4 standard is achieved by removing the redundancy or similarities between the neighbouring frames in a video sequence [1, 2]. There are two kinds of redundancies present in video data: spatial and temporal sampling [2, 3, 4]. The characteristics of a natural video scene, as shown in Figure 1, that are relevant for video processing and compression are spatial characteristics, e.g., texture variation within the scene, number and shape of objects, colour, and temporal characteristics, e.g., object motion, changes in illumination, movement of the camera or viewpoint [3].

Temporal quality refers to the number of frames per second. The motion in a scene appears smoother if more frames are transmitted and received within each time interval. Spatial quality can be expressed as the number of pixels in the video frame, which is the product of its horizontal and vertical resolution as shown in Figure 2. Both temporal and spatial resolutions are usually determined before video encoding, preventing any dynamic trade-offs during the encoding process [5].

The H.264/AVC codec is an attractive candidate for wireless applications including video streaming [6], due to the available features for achieving higher compressions efficiency than previous video standards [7]. The H.264/AVC is widely used in real time video transmission [8].

Figure 1: Still image from a video scene.

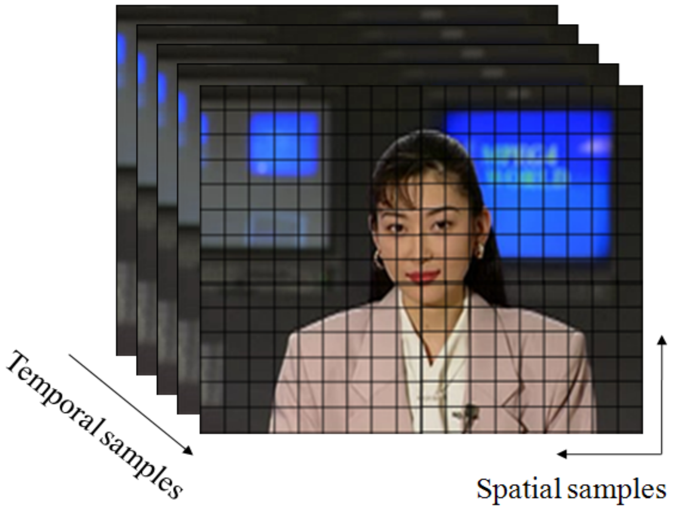


Figure 2: Spatial and temporal sampling of a video sequence.

1.2.2 Video Transmission

Video transmissions over wireless networks are based on two modes, namely downloading and streaming. In download mode, a user downloads the entire video file from the video server and then plays the video file. However, the download mode usually suffers from long and unacceptable transfer time. For streaming mode, all video content does not need to be downloaded before viewing. Instead it is being played when there is an available amount of video frames are being received and decoded by client terminal [9, 10].

Video transmission requires a steady flow of information and delivery of packets by a deadline. However, wireless radio networks have difficulties to provide a reliable service [11]. Video transmission over a dynamic channel, like mobile wireless networks, is much more difficult than over a static channel since the delay and packet loss are not known in advance [12]. Therefore, the wireless networks need some effort in order to use radio resources efficiently [13].

The availability of multiple channels for wireless communication provides an excellent opportunity for performance improvement. The term multiple channels refers to wireless technology that can use more than one radio channel, and multiple channels have been advocated as one approach for enhancing network capacity [14]. The advantages of using of multiple channels over a single channel are to improve performance of limited bandwidth and to increase the channel capacity for video transmission [15].

Multiple-input-multiple-output (MIMO) technology is an effective method to reduce the fading impairments [16] and deliver very high spectral efficiency for the transmission of information over wireless channels [17, 18]. MIMO can be used to transmit the video content over multiple wireless channels. In this case, each path may have lower bandwidth, but the total available bandwidth is higher than the single path.

Multi-path transport can improve the transport reliability by overcoming the instantaneous congestion problem often encountered in the single-path case [19]. Transmitting the video through MIMO is based on spatial multiplexing (SM). SM achieves the increases in throughput with no requirements for additional spectrum. SM relies on transmitting independent data streams from each transmit antenna. The data streams can be multiplexed from the incoming source stream and transmitted through N antennas and are received by the client terminal via N antennas as well. The data can be sent at N-times the rate of a standard terminal [20].

Multiple descriptions coding (MDC) is a source coding technique that generates multiple correlated bitstreams, each of which can be independently encoded and decoded [14]. Each bitstream, called a description, is transmitted through the networks and is expected to follow a different path to reach the destination. The errors that could occur among descriptions are independent and, hence, the performance of MDC is maximal [21, 22]. Therefore, MDC is considered as a promising technique to enhance the error resilience of a video transport system [23] by transmitting the video over multiple independent channels like MIMO, as shown in Figure 3.

Figure 3: Transmitter and receiver structure.

1.2.3 Transmission Protocol

Communication between terminals to transmit a certain amount of data is based on an agreement and a set of rules. Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) are the most used network protocols to communicate between terminals. TCP is connection-oriented, reliable, with flow control, and uses an acknowledgement and retransmission scheme to guarantee delivery of data. In contrast, UDP provides an unreliable and connectionless communication service [24, 25].

Most real-time video services employ UDP as a transport protocol [26]. Compared to TCP, UDP does not involve any retransmission mechanism, which makes it attractive to delay sensitive applications [1, 25, 27]. Further, UDP does not perform any error recovery. Streaming video over wireless networks can be unpredictable, and therefore, UDP is considered as a good candidate for real time video streaming [1, 26].

Real-time Transport Protocol (RTP) provides end-to-end delivery services for data with real-time requirements. Those services include payload type identification, sequence numbering, timestamp, and delivery monitoring. RTP does not provide any mechanism to ensure timely delivery or provide other QoS guarantees, but it will rely on lower-layer services [25].

Video streaming typically run RTP on the top of UDP [1, 25, 28] to make use of its multiplexing and checksum services. Both protocols contribute parts of the transport protocol functionality [25].

1.2.4 Quality of Service

Quality of service (QoS) refers to the ability to provide a satisfactory service during a communication session [29]. Consistently anticipating and meeting users' QoS needs are what distinguishes successful communication services and product providers from their competition [29]. The main concern of video QoS is to provide a satisfactory video quality delivered to the mobile users. Several parameters could affect the quality of video transmission over wireless networks, e.g.,

- **Compression parameters.** The main issues that make video streaming difficult are the large amount of video frames that needs to be transmitted over wireless networks. However, video streaming is compressed in a lossy manner by H.264 codec, leading to smaller compressed representations of the video data than are available with lossless data compression. Therefore, compression plays an important role in video streaming [6, 9, 30, 31]. The nature of the video scene, like the amount of motion, colour, contrast, frame size, and the number of frames transmitted per seconds, can also have an impact on the human perception of the video quality [3, 9, 30, 31].
- **Wireless network parameters.** The main issue with video streaming applications is that it is difficult to achieve guaranteed end-to-end QoS over the entire streaming process [8]. The current best-effort network does not offer any QoS guarantees to video transmission over the networks [9]. Wireless networks performance describes the requirements that must be guaranteed, such as bandwidth, end-to-end delay, and jitter. The network services depend on traffic behaviour and perform according to traffic parameters such as peak data rate or burst loss [1, 9, 12]. Further, the frame loss due to a high error rate during the transmission has a disproportionately negative effect on video quality.

1.3 Aims, Objectives, and Research Questions

The main objective of this thesis is to improve the user perceived quality for real time video streaming over unreliable wireless networks. In this context, the following research questions are addressed.

Primary research question: How can we provide a smooth video playback with satisfactory video quality to the mobile viewer in the presence of network packet loss?

All the research papers in this thesis deal with different aspect of this question. The primary research question been divided into several secondary research questions.

Research question one: How to provide a smooth playback video in the case of low bandwidth and high frame loss rates?

The question is discussed in chapter two (paper 1), where the video frames first will be translated to gray scale to reduce the size of the video and then to be transmitted as duplicated frames over two channels. In case of any losses to the video frames in any channel a switching between channels will be applied to recover the missing frames.

Research question two: How can we improve the user perceived real time video streaming quality by exploiting space-division techniques on the video frames?

This question is discussed in chapter three (paper 2), chapter four (paper 3), and chapter five (paper 4). In paper 2, the frames from the video sequence will be split into two sub-frames, where one sub-frame contains the even pixels and another contains the odd pixels, and then are transmitted over two channels. The missing sub-frame from any channel will be reconstructed from the received sub-frames.

In paper 3, which is an extension of paper 2, the frames from the video sequence are split into four sub-frames. Each sub-frame contains one fourth of the main frame pixels, and each sub-frame are transmitted over different independent channels. The missing of one, two or three sub-frames will be reconstructed from the available sub-frames in the mobile device.

In paper 4, which also is an extension of paper 2, a sub-frame crossing mechanism is considered based on a frame splitting mechanism. Each video frame is split into two sub-frames, and then each sub-frame is combined with another sub-frame from a different sequence position and transmitted as a single frame over a single wireless channel. In case of frame losses, there is still a possibility that one of the sub-frames will be received by the mobile device. The missing sub-frame can then be reconstructed. A rate adaptation mechanism will be also highlight in this work. We show that the server can skip up to 50% of the sub-frames and we can still be able to reconstruct the receiving sub-frame and play it on the mobile device.

Research question three: How can we improve the user perceived real time video streaming quality by exploiting time interleaving techniques on the video frames?

The question is discussed in chapter six (paper 5). The sequence of the video frames will be reordered as a group of even followed by odd frames to eliminate the frozen pictures by avoiding a sequence of neighbouring frames to be lost and allowed at least two neighbour frames to be present in the mobile device based on the propose Time Interleaving Robust Streaming (TIRS) technique.

The TIRS will be applied before the video been encoded and compressed by the H.264 encoder. The TIRS can be implemented for different interleaving time to distribute the big loss of frames on different positions in the streaming sequence and to avoid a sequence of interruption to the streaming video.

1.4 Research Methodology

The simulation environment used in this thesis to simulate the proposed techniques is Simulink [32]. Simulink is used according to the available features that allow us to stream the video in real time. The output test video is evaluated by using the average subjective quality of a video sequence across a number of human viewers, known as the Mean Opinion Score (MOS).

International Telecommunication Union Recommendations (ITU-R) for subjective video quality assessment include specifications for how to perform many different types of subjective tests. The most commonly used one for video quality evaluation is the MOS that been recommended by the ITU [33]. MOS is a subjective quality metric obtained from a panel of human observers. It has been regarded for many years as the most reliable form of quality measurement technique [34]. MOS is very tedious, expensive, and time consuming [35]. The subjects provide the perception of quality on a continuous linear scale which was divided into five equal regions marked with the MOS-scale adjectives described in Table 1.

The natural way to compare the quality of two video streams is to evaluate it by using human observers. Typically, an observer examines a set of videos in a controlled environment and assigns a numerical score to each of the videos. Each video's scores are recorded and averaged later as its MOS.

Table 1: Subjective quality rating scales [33].

Rating	MOS	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

The experimental study is carried out for different frame losses condition and for different types of videos, where the chosen videos have different motion levels. The sample size for each study in this thesis is 30 persons where the output results will be based on their observation. The experiments are taken place in the university library and the physical environments are set according to the ITU-R recommendations, where each person evaluates and judges the video based on 5-score scale as shown in Table 1. The people who are involved in the study should not have any professional backgrounds that are related to video. The observers should concentrate on the differences in the same video. Each video is played two times to the observer. The first video is played with the content of the effect of frame losses where some part of the video is frozen, while the second video is played based on the development for our proposed techniques. The measurement method is applied to all the videos played for the same observer. The evaluation form should be marked by the observer for each time the video been played.

The open source ffmpeg codec software [36] is used as well in this thesis to compress the video. The reason of using ffmpeg codec is to study the effect of the proposed technique on the video size. The compressed video sizes are used to compare between the original video and to the proposed techniques video.

1.5 Thesis Contributions

In this thesis, we propose several techniques for streaming video over unreliable networks. The proposed techniques can adapt to the current network conditions and requirements, and possibly could help to overcome the effect of bandwidth variation and outage interruption time on the video streaming. The main contributions of this work are:

Chapter two: We proposed a mechanism to play the complete video frame sequence in the mobile device over low bandwidth channels. This is done by converting the video frames in the server from colour to gray scale. The main idea of colour conversion is to reduce the video frames size. The second stage is to create duplicated frames that can be transmitted over two channels in the cellular network. After the two video streams have been received by the mobile device, a switching mechanism between the two streams will take place whenever there is a missing frame from one of the video sequences. This will help to improve the robustness of the video transport, since a lost frame in one channel can be recovered from other channel.

Chapter three: A novel frame splitting technique to create two sub-frames out of each frame, where one sub-frame contains the even pixels and the another contains the odd pixels. Each sub-frame will be queued in two different buffers, and will be streamed over two wireless channels. The second stream will be delayed for 2 seconds after the first stream. The reason behind that is to minimize the effects to the same sub-frames of any dropping frames or interruption to the wireless channel under different networks condition. Streaming the video based on independent transmission on the two channels will be used in order to achieve such purpose. If there is a missing sub-frame from any subsequence during the transmission a reconstruction mechanism will be applied in the mobile device to recreate the missing sub-frames.

Chapter four: An extension of the technique in chapter three where each frame is split into four sub-frames based on a pixel distribution, and each sub-frame contains one fourth of the main frame pixels. Each sub-frame will represent its own subsequence, the first subsequence is transmitted without any delay, and the second subsequence will be delayed for 0.5 second the third subsequence will be delayed for 1.0 second, while the fourth subsequence will delayed for 1.5 seconds. The reason for implementing the subsequence transmission delay is to minimize the effect of any dropping or corruption to the sub-frames that belong to the same frame over a wireless channels and under different networks condition.

Chapter five: A technique to split each video frame into two sub-frames. Each sub-frame will then be combined with another sub-frame from a different sequence position and transmitted as a single frame over a single wireless channel. In case of frame losses or frame corruption from the streaming video, there is still a possibility that one of the sub-frames will be received by the mobile device. The received sub-frames will be reconstructed and played on the mobile screen.

Chapter six: A novel interleaved approach is proposed to reorder the video frames before streamed over wireless networks. The frames will be grouped as sequences of even frames followed odd frames and then streamed to the mobile device. Streaming the video frames in different order by using the proposed Time Interleaving Robust Streaming (TIRS) is to eliminate the frozen picture by avoiding a sequence of neighbouring frames to be lost and allowed at least the two surrounding frames to be present in the mobile device. If there is a missing frame in the mobile device a reconstruction mechanism is applied.

1.6 Research Validity

The results that are obtained from the system design and the subjective evaluation are quite encouraging, but there are some limitations that are needed to be highlighted.

Regarding the simulation study, we design our techniques in the way that allow us to stream the videos from the server node to the clients node without considering the communication setup time between the server and the client, which we assume it constant in all our studies. Further, there are different types of failures that could occur during the streaming video, e.g., error transmission and intermediate node failure, which could have different effects on the streaming video.

During the experimental study, some of the user panels requested to re-play the videos as the videos length is short, and they feel that, they need to play the videos again for fair judgments. Miras [37] also highlight that the video sequence is not long enough to experience different kinds of impairments that could occur to the videos test sequence.

1.7 Related Work

The use of multichannels through the transport network for streaming has been proposed to overcome the frame losses and delay problems that afflict the streaming video and low latency communication. In addition, it has long been known that multichannels can improve fault tolerance and link recovery for frame delivery, as well as provide larger aggregate bandwidth, load balancing, and faster bulk data downloads [38].

Apostolopoulos [38] suggest to use two different channels to transmit even and odd frames that will be encoded by using Multiple Description Coding (MDC), he suggests that it can be beneficial to transmit different amounts of traffic on different channels. It may consist of two separate encoders and decoders that alternate previous decoded frame it uses to perform the prediction. If there are no errors or frame losses and both even and odd streams are received correctly, then both streams are decoded to produce the full frames sequence for final display. If one stream has an error or frame losses then the state for that stream is incorrect and there will be error propagation for that stream. However, the other independently stream can still be accurately and straightforwardly decoded to produce usable video.

Lin et al. [39], investigate the influence of two important characteristics on the performance of the diversity level of the space time code and the error resilient ability of the video coder. For MIMO video system, the selection of diversity levels and the selection of an error-resilient video coder with appropriate redundancy are not independent of each other. Where the video data need to be lossy coded before transmission over three transmitting and receiving antennas. The distortion caused by the lossy source encoder directly depends on the availability of the bit rates. Although a space time coder with full diversity can provide strongest error protection of the source data, it usually supports a lower bit rate, which means that the video encoder can only output low-quality pictures. On the other hand, a space time coder with maximum rate (minimum diversity) allows the highest bit rate for video coding, but its weak error correction ability may cause devastating effects in the received video quality. The results show that the appropriate source coding and channel coding schemes depends on the channel environment. Hence, to obtain the best overall system performance, the source coding and channel coding parameters need to be chosen jointly and with regard to the channel quality.

Tesanovic et al. [40], proposed a new scheme for video transmission for efficient and robust video streaming video over wireless channels, through a combination of MIMO and MDC technology. Two complementary MIMO techniques are used: space time block coding and spatial multiplexing. The quality of the reconstructed video, already enhanced by the inherent MIMO systems' properties, is further improved through the use of MDC. MDC has evolved and has been adapted to lossy packet networks. With correct design, MDC can exploit the interactions between descriptions when losses occur in multichannels wireless communications to reliably recover to the video.

While another researcher like Claypool and Zhu have proposed different direction for avoiding the effects of channel error and frame losses during the transmission. Claypool and Zhu [41] present an interleaving algorithm on MPEG encoded video frames. Interleaving represents a good technique for wireless networks which spreading the losses over the video stream. The interleaving to the sequence of frames is been applied before the MPEG encoder for interleaving distance algorithm for two and five to the video frames sequence.

1.8 Conclusion

The main contribution of this thesis is to improve the end user perceived quality to the real time video that is streamed over unreliable networks. The transmission rate of the channel varies from time to time and deplaned on the available bandwidth, while the networks are unable to control the amount of video frames that are transmitted over the wireless channels. The video frames could be lost, delayed, or could be affected by errors and become unreadable by the decoder, therefore the video streaming may experience frame losses that could causes degradation of the quality of video for several seconds, and the mobile users will notice a set of frozen pictures on their screen.

Therefore, we proposed several techniques to overcome the effect of the network interruption on the end video quality, where the techniques are designed in the way that provides addition delay time to the initial setup time for the video transmission as a cost of our proposed techniques. The users will face extra delay time in seconds as they will get a smooth play video in their mobile devices until if the network lost frame rate is high.

References

[1] J. G. Apostolopoulos, W.-T. Tan, and S. J. Wee, “Video Streaming: Concepts, Algorithms, and Systems,” HP Technical Reports, HPL-2002-260, 2002.

[2] W.T. Staehler, and A. A. Susin, “Real-Time 4x4 Intraframe Prediction Architecture for a H.264 Decoder,” International Telecommunications Symposium ITS‘06, pp. 416 - 421, 2006.

[3] I. E. G. Richardson, “H.264 and MPEG-4 Video Compression Video Coding for Next-Generation Multimedia,” John Wiley & Sons Ltd, 2003.

[4] M. Ezhilarasan, P. Thambidurai, V. Rajalakshmi, R. Ramya and M. Vishnupriya, “An Improved Transformation Technique for H.264/Advanced Video Coding,” International Conference on Computational Intelligence and Multimedia Applications, ICCIMA’07, pp. 123-127, 2007.

[5] Y. Lee, H. Lee, H. Shin, “Adaptive Spatial Resolution Control Scheme for Mobile Video Applications,” IEEE International Symposium on Signal Processing and Information Technology, pp. 977-982, 2007.

[6] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, “H.264/AVC in Wireless Environments,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 657-673, 2003.

[7] J. Zhang, F. Liu, H. Shao, G. Wang, “An Effectives Error Concealment Framework for H.264 Decoder Based on Video Scene Change Detection,” 4th International Conference on Image and Graphics, ICIG’07, pp. 285-290, 2007.

[8] X. Zhou, and C-C. Jay Kuo, “Robust Streaming of Offline Coded H.264/AVC Video Via Alternative Macroblock Coding,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, No. 4, pp. 425-438, 2008.

[9] D. Wu, Y. T. Hou, W. Zhu, Y-Q. Zhang, and J. M. Peha, “ Streaming Video over the Internet: Approaches and Directions, ” IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Streaming Video, Vol. 11, No. 3, pp. 282-300, 2001.

[10] D. N. Sujatha, K. Girish, and K. V. Rajesh, K. R. Venugopal, L. M. Patnaik, “ Preemption-Based Buffer Allocation in Video-on-Demand System,” International Conference on Advanced Computing and Communications, ADCOM ’06, pp. 523- 528, 2006.

[11] X. Zhu, and B. Girod, “ Video Streaming over Wireless Networks,” In Proceedings of European Signal Processing Conference, EUSIPCO`07, pp. 1462-1466, 2007.

[12] V. Vassiliou, A. Pavlos, G. Iraklis, P. Andreas, “ Requirements for the Transmission of Streaming Video in Mobile Wireless Networks,” International Conference on Artificial Neural Networks, ICANN’06, pp. 528-537, 2006.

[13] J. Y. Kim, and D-H. Cho, “Enhanced Adaptive Modulation and Coding Schemes Based on Multiple Channel Reporting for Wireless Multicast Systems,” VTC ‘05, pp. 725 - 729, 2005.

[14] J. Cao and C. Williamson, “Towards Stadium-Scale Wireless Media Streaming,” In Proceedings of IEEE/ACM MASCOTS’06, Monterey, California, pp. 33-42, 2006.

[15] C. Chereddi, P. Kyasanur, and, N. H. Vaidya, “Design and Implementation of a Multi-Channel Multi-Interface Network,” In Proceedings of the 2nd International Workshop on Multi-Hop ad Hoc Networks: from Theory to Reality. REALMAN ´06, pp. 23 – 30, 2006.

[16] Y. Liu, L. Sun, and S. Yang, “A Virtual MIMO Transmission Scheme in Wireless Video Sensor Network,” International Symposium on Computer Network and Multimedia Technology, CNMT ‘09, pp. 1-4, 2009.

[17] L. Zheng, and D. N. C. Tse, “Diversity and Multiplexing: A Fundamental Tradeoffs in Multiple-Antenna Channels,” IEEE Trans. Info. Theory, Vol. 49, No. 5, pp.1073-1096, 2003.

[18] L. Dong, H. Ling, and R. W. Heath, “ Multiple-Input Multiple-Output Wireless Communication Systems using Antenna Pattern Diversity,” In Proceedings of the IEEE Global Telecommunications Conference, GLOBECOM’02, pp. 997-1001, 2002.

[19] J. Zhou, H-R. Shao, C. Shen, and M-T Sun, “Multi-Path Transport of FGS Video,” April, 2003. Technical report. http://www.merl.com.

[20] M. Tesanovic, D.R. Bull, and A. Doufexi, “Enhanced Error-Resilient Video Transport over MIMO Systems using Multiple Descriptions,” Vehicular Technology Conference, pp. 1-5, 2006.

[21] J. Kim, “ Layered Multiple Description Coding For Robust Video Transmission over Wireless Ad-Hoc Networks,” World Academy of Science, Engineering and Technology, pp. 163- 166, 2006.

[22] Y. Lee, Y. Altunbasak, and R. M. Mersereau, “Optimal Packet Scheduling for Multiple Description Coded Video Transmissions over Lossy Networks,” IEEE Global Telecommunications Conference, Globecom’ 03, 2003.

[23] Y. Wang, A. R. Reibman, and S. Lin, “ Multiple Description Coding for Video Delivery,” In the proceedings of IEEE, Vol. 93, No. 1, pp. 57-70, 2005.

[24] A. Löfgren, L. Lodesten, S. Sjöholm, H. Hansson, “ An Analysis of FPGA-based UDP/IP Stack Parallelism for Embedded Ethernet Connectivity,” In Proceedings of the 23rd IEEE NORCHIP Conference, pp.94-97, 2008.

[25] L. Xu, and S. Ai, “ A New Feedback Control Strategy of Video Transmission Based on RTP,” 1st IEEE Conference on Industrial Electronics and Applications, 2006.

[26] G. Ding, H. Ghafoor and B. Bhargava, “Error Resilient Video Transmission over Wireless Networks,” In Proceedings of the 6th IEEE International Symposium on Object-oriented Real-time Distributed Computing, 2003.

[27] H. Zheng, and J. Boyce, “An Improved UDP Protocol for Video Transmission over Internet-to-Wireless Networks,” IEEE Transactions on Multimedia, Vol. 3, No. 3, pp. 356-365, 2001.

[28] A. Alexiou, C. Bouras, V. Igglesis, “A Decision Feedback Scheme for Multimedia Transmission Over 3G Mobile Networks,” The Second IEEE and IFIP International Conference on Wireless and Optical Communications Networks, WOCN‘05, pp. 357 – 361, 2005.

[29] S. Fredricsson, and C. Perey, “Quality of Service for Multimedia Communications”. http://www.tfi.com/pubs/ntq/articles/view/95Q4_A8.pdf

[30] G. J. Sullivan, P. Topiwala, and A. Luthra, “The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions,” The SPIE Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard: H.264/AVC, pp 1-22, 2004.

[31] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “ Overview of the H.264/AVC Video Coding Standard,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 560-576, 2003.

[32] www.mathworks.com.

[33] International Telecommunication Union. Methodology for the Subjective Assessment of the Quality of Television Pictures. ITU-R, Rec. BT.500-11, 2002.

[34] M.Martinez-Rach, O.López, P.Piñol, M.P. Malumbres, J. Oliver, Carlos T. Calafate, “ Quality Assessment Metrics vs. PSNR under Packet Loss Scenarios in MANET Wireless Networks,” International Workshop on Mobile Video, MV ‘07, 2007.

[35] Z. Wang, and A.C. Bovik, “A Human Visual System-Based Objective Video Distortion Measurement System,” International Conference on Multimedia Processing and Systems, ICMPS’00, 2000.

[36] www.ffmpeg.org.

[37] D. Miras, “A Survey of Network QoS Needs of Advanced Internet Applications,” in Internet2 QoS Working Group, 2002.

[38] J. G. Apostolopoulos, “Reliable Video Communication over Lose Packet Networks using Multiple State Encoding and Path Diversity,” Visual Communication and Image Processing Conference, pp. 392-409, 2001.

[39] S. Lin, A. Stefanov, and Y. Wang, “On the Performance of Space–Time Block-Coded MIMO Video Communications,” IEEE Transactions on Vehicular Technology, Vol. 56, No. 3, pp. 1223-1229, 2007.

[40] M. Tesanovic, D. R. Bull, and A. Doufexi, “Enhanced Error-Resilient Video Transport over MIMO Systems using Multiple Descriptions,” IEEE Vehicular Technology Conference, pp. 1-5, 2006.

[41] M. Claypool and Y. Zhu, “Using Interleaving to Ameliorate the Effects of Packet Loss in a Video Stream,” In Proceedings of the International Workshop on Multimedia Network Systems and Applications (MNSA), 2003.

CHAPTER TWO

Graceful Degradation of Mobile Video Quality over Wireless Network

Hussein Muzahim Aziz and Lars Lundberg

Abstract

Real-time video transmission over wireless channels has become an important topic in wireless communication because of the limited bandwidth of wireless network that should handle high amount of video frames. Video frames must arrive at the client before the playout time with enough time to display the contents of the frames. Real-time video transmission is particularly sensitive to delay as it has a strict bounded end-to-end delay constraint; video applications impose stringent requirements on communication parameters, such as frame lost and frame dropped due to excessive delay are the primary factors affecting the user-perceived quality. In this study we investigate ways of obtaining a graceful and controlled degradation of the quality, by introducing redundancy in the frame sequence and compensating this by limiting colour coding and resolution. The effect of that is to use double streaming mechanism, in this way we will obtain less freezing at the expense of limited colours and resolution. Our experiments, applied to scenarios where users can observe three types of dropping load for real time video streaming, the analytical measurements tools are used in this study to evaluate the video quality is the mean opinion score and we will demonstrate this and argue that the proposed technique improves the use perceived of the video quality.

Keywords

Streaming video, duplicate frames, multichannel, dropping rate, switching stream and MOS.

2.1 Introduction

Mobile wireless network becomes very popular now due to the wide spread of computer laptops, mobile devices and PDAs. Mobile wireless networks are expected to support different type of services, such as video services which make a great demand on the wireless networks bandwidth. Bandwidth is one of the most critical resources in wireless networks, and thus, the available bandwidth of wireless networks should be managed in an efficient manner (Chang & Chen, 2003). The video coding standards are being developed to satisfy the requirements of applications for various purposes, excellent video/picture quality, higher coding efficiency, and error robustness (Zhang, 2005).

Song et al (Song et al, 2002) identifies the major difference between mobile wireless environments and wired one as follows: (i) the available bandwidth and (ii) the size of screen. They believe that the impact of two compression parameters may change dependent upon the maximum screen size, and the available bandwidth. Low bandwidth provided by wireless networks is often insufficient to support streaming video over a single channel; multiple channels can be used to provide a logical channel with a sufficient (or nearly sufficient) bandwidth (Xu & Hemami, 2002). Video communication over a dynamic environment, like a mobile wireless network is much more difficult than over a static channel, since the bandwidth, delay, and packet loss are not known in advance and are unbounded (Vassiliou et al, 2006).

Video applications involve real-time streams of information that need to be transported in a dynamic and high-performance way. The goal of this paper is to come up with a mechanism to play the complete video frame sequence in the mobile station over error prone channels, this can be done by transferring the video frames in the server from colour to gray scale, the main idea of colure transfer is to reduce the video frames size while the second stage is to create a duplicated frames that can be transmitted over multiple channels in the cellular network. After the two video streams have been received by the mobile client, the switching mechanism between the two streams will take place whenever there is a missing frame from the video sequence. This will improve the robustness of the video transport, since a lost frame in one channel can be recovered from other channel. To the best of our knowledge, our work is the first study on playing the complete video frame sequence till if there is a frame lost at any stream.

2.2 Background and Related Work

Streaming video is the classical technique for achieving real time video directly over the network without downloading the entire file before playing the video (Bai & Williamson, 2004), (Xiaozhen et al, 2005) and (Zhu et al, 2003). Streaming video requires high reliability with a low bounded jitter (i.e. variation of delays) and reasonably high transmission rate (Hsu et al, 1999). Video streaming requires a steady flow of information and delivery of frames by a deadline; wireless radio networks have difficulties to provide such reliably services (Zhu & Girod, 2007). The availability of multiple channels for wireless communication provides an excellent opportunity for performance improvement. The term multichannel refers to wireless technology that can use more than one radio channel. The use of multiple wireless channels has been advocated as one approach for enhancing network capacity. Some wireless devices achieve this property using multi-radio systems, with each interface communicating on a different physical channel. Other devices have just a single radio transceiver, which is tuneable to any of the available channel (Cao & Williamson, 2006). The use of multiple paths through the transport network for streaming has been proposed to help overcome the loss and delay problems that afflict streaming media and low latency communication. In addition, it has long been known that multiple paths can improve fault tolerance and link recovery for data delivery, as well as provide larger aggregate bandwidth, load balancing, and faster bulk data downloads (Apostolopoulos, 2001). Zhou et al (Zhou et al, 2003), present a transmission scheme to improve MPEG-4 streaming using multipath. The MPEG stream is divided into a base and enhancement layer. The base layer is duplicated over the paths available to provide robustness for transmission of important frames. The enhancement layer content is then separated into multiple descriptions and sent over the multiple paths leading to incremental increase in the video quality with reception of more and more enhancement packets. While Apostolopoulos (Apostolopoulos, 2001) use two different paths to send even and odd frames encoded using Multiple Description Coding (MDC), he suggests that it can be beneficial to send different amounts of traffic on different paths. Most of the schemes discussed above are verified for two paths. With two paths all the above algorithms provide performance improvements over the single path but none of them manage to deliver the complete video frame sequence to the mobile station as been transmitted by the server.

2.3 Streaming Video over Two Channels

Mobile video streaming is characterized by low resolutions and low bit rates. The bit rates are limited by the capacity of UMTS (Universal Mobile Telecommunications System) radio bearer and restricted processing power of mobile terminals. The commonly used resolutions are Quarter Common Intermediate Format (QCIF, 176×144 pixels) for cell phones (Ries et al, 2007). Wireless network have a limited bandwidth which have not been able to handle the continues of the video frame sequence and also with the high possibility that the video frame could be timed out while they are waiting in the buffer, which all of what we described above could affect the video quality for the mobile viewer. Shenoy and Vin (Shenoy & Vin, 1995), suggested that the video server can partition each video stream into two sub-streams (a low-resolution and a residual component stream) in order to support interactivity. During the interactive mode, only the low-resolution stream is transmitted to the client, this can reduces the amount of data that needs to be retrieved and sent to the client's mobile. While, multichannel is been proposed recently in mobile cellular network by many researcher like (Zheng et al, 2005) and (Chereddi &Vaidya, 2006), when multichannel provides information to clients via multiple channels.

During the busy traffic, the mobile station request from the server to establish the connection to stream the video, the corresponding video streams are obtained by encoding the original uncompressed video frame from colure to gray scale; this will reduce the frame size by 34%. In the mean time a duplicated video frame will be generated by the server and will be queued in two different buffers and it will be ready to stream the video in two wireless channels but the second stream will be delayed by 2 seconds after the first stream as in Figure 1, the reason behind that is to avoid dropping the same frames from each stream, while some frames are timed out under different network load condition.

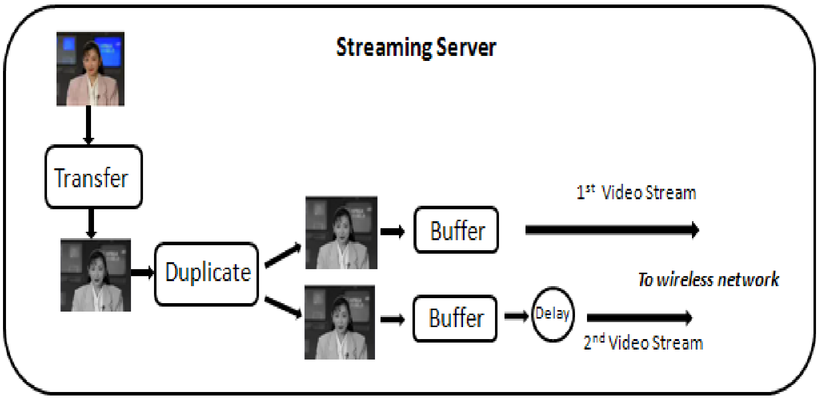


Figure 1: Streaming duplicate frames over two wireless channels.

When the mobile client establish the connection with video server, the client will start receiving the stream of video and it will be held in the jitter buffer until the available amount of frames have been received to start playing, this is for the normal case when there is one channel handling a single stream, while according to our proposed system after the first stream has been received by the mobile device it will be delayed for 2 seconds until the mobile station start receiving the second stream of video frames. After both streams received the right amount of frames to be played, the mobile's user will start viewing the video frame sequence from any healthy stream.

In case, frames are dropped in one stream, the switching mechanism will be applied between the two streams at the time when the dropping occurs. Switching between streams are used to make sure that the mobile screen will display the complete video frame sequences in case there is a missing frame from any stream channel has been shown in Figure 2. If the same frame is lost in both streams, then we cannot handle this and there will be some freezing in the video.

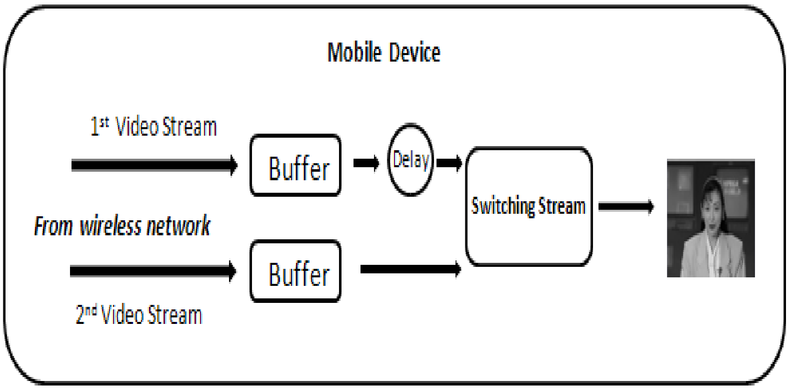


Figure 2: Receiving duplicate streams of video in the same mobile device.

The simulation environments used for this study is Simulink (www.mathworks.com). Three different dropping rates for the same network traffic have been run for the proposed work, to observe the switching mechanism. Under the light traffic where the frame dropped rate is between 3-4%. While for the medium traffic load the dropped frame rate is between 6-7% and for the high traffic load the dropped frame rate is between 8-9%. The sequence of video frames which are received by the mobile buffer could not contains the same sequences of frames as some of them could be dropped during the transmission and also the first stream of video need to wait for 2 seconds until the next stream arrived. Two buffers have been introduced in this work to handle the arriving frame sequence in each stream, and also to hold the stream which is not been played until the switching between streams has taken place.

2.4 Subjective Viewing Test

2.4.1 Testing Materials and Environments

The video test sequences used in this work were the samples of video sequences Akiyo, Foreman and News, the video sequences were chosen because of their deferent characteristics. the number of frames are coded for each video are 25 frames/second with a resolution of 176 x 144, while the transmission rate are 30 frames/second and the number of frames are transmitted are 1800 frames. A 17 FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of 1680 x 1050 pixels, while the video sequences for the original and our proposed scenario are displayed with resolution of 176 x 144 pixels in the center of the screen with black background with a duration of 60 seconds for each video sequence.

2.4.2 Testing Methods

According to the guidelines outlined in recommendation BT.500-11 of the radio communication sector of the International Telecommunication Union (ITU-R) (International Telecommunication Union, 2002), a subjective experiment was conducted for two groups of viewer's. Both experiments have been conducted at Blekinge Institute of Technology in Sweden. Two different groups observe two different scenarios, the first group, observed the first scenario where the evaluation done for normal video transmission over wireless networks with the original video colure display and for three different load effect, while the second group, observed our proposed scenario (the second scenario). Thirty non-expert test subjects, 28 males and 2 females, participated in each group. They were all university staff and students and their ages range of 20 to 41 of age.

The score grades in this methods range from 0 to 100 which is mapped to the quality ratings were made on the 5-grade discrete category scale labelled with Excellent, Good, Fair, Poor and Bad. We conducted a physical experiment in a Lab, with controlled lighting and set-up conforming to the ITU-R recommendation. The amount of data gathered from the two subjective experiments groups with respect to the opinion scores that were given by the individual viewers. A concise representation of this data can be achieved by calculating conventional statistics such as the mean score and confidence interval, of the related distribution of scores. The statistical analysis of the data from the subjective experiments reflects the fact that perceived quality is a subjective measure and hence may be described statistically. The mean opinion score (MOS), obtained as the arithmetic mean of the scores, is used as the degradation assessment result as (1)

$$\mu = \frac{1}{N}\sum_{i=0}^{\mathrm{N}} \mathrm{u_i} \qquad (1)$$

Where u_i denotes the opinion score given by i^{th} viewer and N is the number of viewers. The confidence interval associated with the MOS of each examined video is given by

$$[\,\mu - \delta\,, \mu + \delta\,] \qquad (2)$$

Its note that the deviation term δ in (2) can be derived from the standard deviation σ and the number N of viewers and given for 95% confidence interval according to ITU-R recommendation

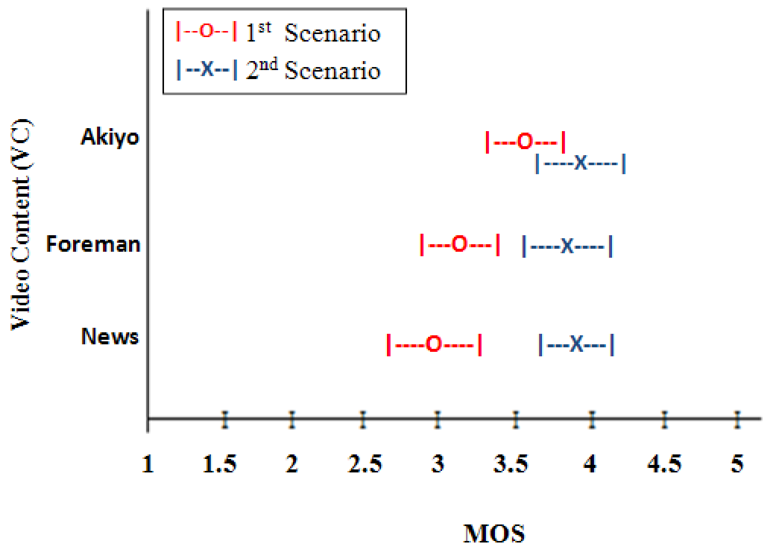
$$\delta = 1.96\frac{\sigma}{\sqrt{N}} \qquad (3)$$

Where the standard deviation σ , is define as square root of the variance

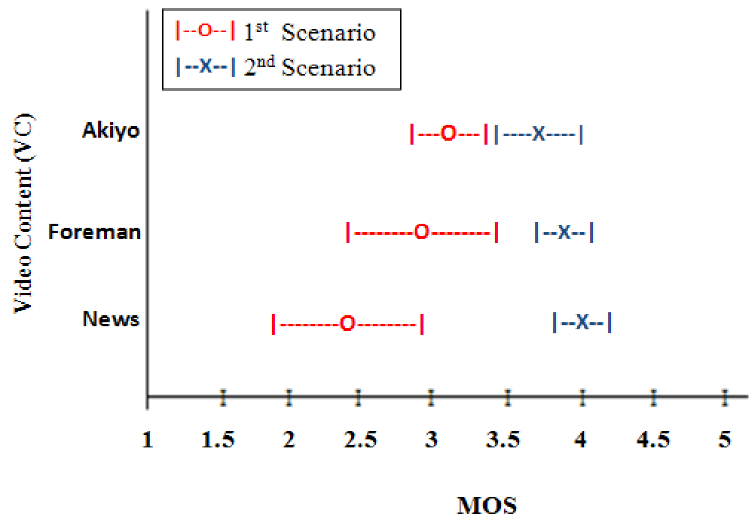
$$\sigma^2 = \sum_{i=1}^{N} \frac{(\mu - u_i)^2}{N - 1} \qquad (4)$$

2.4.3 Testing Results and Discussions

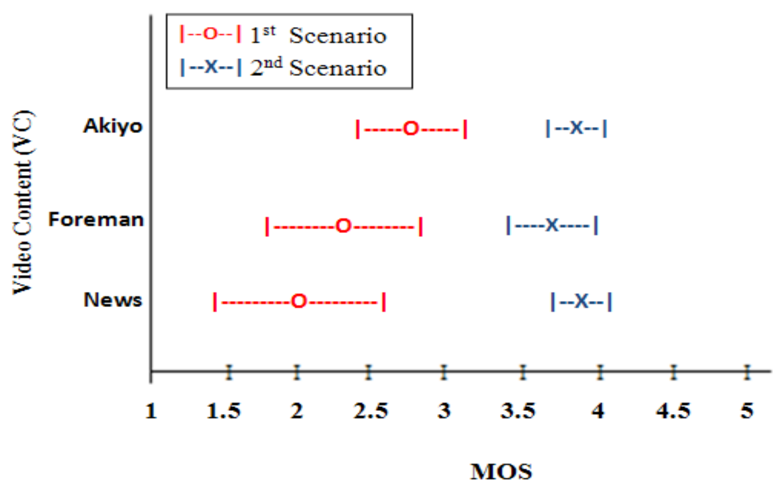
It is well known that peak signal-to-noise ratio (PSNR) does not always rank quality of an image or video sequence in the same way that a human being. There are many other factors considered by the human visual system and the brain (Martinez-Rach et al, 2007). The mean opinion score (MOS) measurements are used to evaluate the video quality. Quality of video is subject of the personal opinion; this means that the quality of service improvements for video transmission has the only goal to satisfy the average human watching the contents of the video. The MOS is usually obtained through human evaluation tests; two groups of 30 students/lecturers are conduct for these measurements. Figure 3, shows the comparison test for video content (VC) and for different dropping rate percentage, where the center and span of each horizontal bar indicate the mean score and the 95% confidence interval, respectively.



a. Light dropped rate.



b. Medium dropped rate.



c. High dropped rate.

Figure 3: The MOS for different video contents via different dropping rate.

For the first scenario with different dropping rate, its shows that, the MOS lower than 4 corresponding to the ‘bad’, ’poor’, ’fair’, and ‘good’ ranks of the five-level quality scale whereas for the second scenario with different load, the MOS larger than 3, corresponding to the ‘good’ and ‘excellent’ ranks. It can be observed that the material present to the viewers resulted in a wide range of perceptual quality ratings indeed for both experiments. After we analysis their score we feel that our propose scenario (2nd scenario) could be a perfect technique to make sure that the all video frame sequence played smoothly and without any freezing during the transmission.

2.5 Conclusion

Real-time video streaming is particularly sensitive to delay, frame loss and frames dropped due to excessive delay are the primary factors that have a negative effect on the received video quality over wireless network. Every video frame must arrive at the client before its playout time, with enough time to decode and display the contents of the frame. If the video frame does not arrive on time, the playout process will pause and the frame is effectively lost. Our proposed technique is to run the complete frame sequence in the mobile device, this can been done by streaming the same gray video two times and transmitted over two wireless channels and can be received by two buffers in the same mobile device.

A switching mechanism is highly needed for video transmission over multichannel in the mobile device and for two reasons: (1) to prevent video frames from missing deadline, (2) dropping or corrupted frames in the transmission channels, this can be done by switching between video streaming channels and it will provide the right sequence of the frame to be played in the mobile devices. Through the simulation study with human opinion it showed that there is a significant performance improvement for video quality under different drooping load over wireless network, we conclude that our proposed technique is useful to provide a smooth video to the mobile viewer. This particularly true for mobile network with poor capacity and high dropping rates.

Acknowledgments

We would like to thanks the staff and students of the School of Engineering at the Blekinge Institute of Technology, Ronneby, Sweden for participating in the subjective experiments.

References

Apostolopoulos J. G., 2001. Reliable Video Communication over Lose Packet Networks using Multiple State Encoding and Path Diversity,” Visual Communication and Image Processing Conference, pp. 392-409.

Bai G. and Williamson C., 2004. The Effects of Mobility on Wireless Media Streaming Performance. Proceedings of Wireless Networks and Emerging Technologies (WNET), Banff, AB, Canada, pp. 596-601.

Cao J. and Williamson C., 2006. Towards Stadium-Scale Wireless Media Streaming. Proceedings of IEEE/ACM MASCOTS, Monterey, California, pp. 33-42.

Chang J-Y and Chen H-L, 2003. Dynamic-Grouping Bandwidth Reservation Scheme for Multimedia Wireless Networks**.** IEEE Journal on Selected area in Communications, Vol. 21 No. 1, pp. 1566 – 1574**.**

Chereddi C. P. K. & Vaidya N. H., 2006. Design and Implementation of a Multi-Channel Multi-Interface Network. The 2nd International Workshop on Multi-hop Ad Hoc Networks: From Theory to Reality, Florence, Italy, pp. 23 – 30.

Hsu C.Y. et al, 1999. Rate Control for Robust Video Transmission over Burst-Error Wireless Channels. IEEE Journal on Selected Areas in Communications, Vol. 17, No. 5, pp 756 - 773.

International Telecommunication Union, 2002. Methodology for the subjective assessment of the quality of television pictures," ITU-R, Rec. BT.500-11.

Martinez-Rach M. et al, 2007. Quality Assessment Metrics vs. PSNR under Packet Loss Scenarios in MANET Wireless Networks. International Workshop on Mobile Video (MV 2007), Augsburg, Germany.

Ries M. et al, 2007. Performance Evaluation of Mobile Video Quality Estimators. 15th European Signal Processing Conference, Poznan, Poland, pp. 159-163.

Song et al, 2002. Empirical Study of User Perception Behavior for Mobile Streaming. International Multimedia Conference, Multimedia'02, Juan-les-Pins, France, pp. 327 – 330.

Shenoy P. J. and Vin H. M., 1995. Efficient Support for Scan Operations in Video Servers. ACM Multimedia Conference, San Francisco, CA, pp. 131-140.

Vassiliou V. et al, 2006.Requirements for The Transmission of Streaming Video in Mobile Wireless Networks. International Conference on Artificial Neural Networks (ICANN2006), Athens, Greece, pp. 528-537.

Xiaozhen (Jean) C. et al, 2005. Media Streaming Performance in a Portable Wireless Classroom Network. Proceedings of IASTED European Workshop on Internet Multimedia Systems and Applications (EuroIMSA), Grindelwald, Switzerland, pp. 246-252.

Xu W. and Hemami S. S., 2002. Spectrally Efficient Partitioning of MPEG Video Streams for Robust Transmission over Multiple Channels. IEEE Proc. of International Conference on Image Processing, Rochester, NY,USA.

Zhang Y., 2005. Advanced Video Coding: MPEG-4/H.264 and Beyond.CMPT 880 - SPECIAL TOPICS: MULTIMEDIA NETWORKING, 2005. http://www.cs.sfu.ca/~johnnyz/personal/cmpt880-2/ johnnyz 880_ final.pdf

Zheng B. et al, 2005. TOSA: A Near-Optimal Scheduling Algorithm for Multi-Channel Data Broadcast. The 6th International Conference on Mobile Data Management, New York, USA, pp. 29 – 37.

Zhou J. et al, 2003. Multi-path Transport of FGS Video.13th International Workshop on Packet Video, Nantes, France.

Zhu H. et al, 2003. Bandwidth Scalable Source-Channel Coding For Streaming Video over Wireless Access Network**s.** WNCG Wireless Networking Symposium, Austin, USA.

Zhu X. and Girod B., 2007. Video Streaming Over Wireless Networks. Proc. European Signal Processing Conference, (EUSIPCO-07), Poznan, Poland, pp. 1462-1466.

CHAPTER THREE

Eliminating the Freezing Frames for the Mobile User over Unreliable Wireless Networks

Hussein Muzahim Aziz, Håkan Grahn, and Lars Lundberg

Abstract

The main challenge of real time video streaming over a wireless network is to provide good quality service (QoS) to the mobile viewer. However, wireless networks have a limited bandwidth that may not be able to handle the continues video frame sequence and also with the possibility that video frames could be dropped or corrupted during the transmission. This could severely affect the video quality. In this study we come up with a mechanism to eliminate the frozen video and provide a quality satisfactory for the mobile viewer. This can be done by splitting the video frames to sub-frame and transmitted over multiple channels. We will present a subjective test, the Mean Opinion Score (MOS). MOS is used to evaluate our scenarios where the users can observe three levels of frame losses for real time video streaming. The results for our technique significantly improves the indicate perceived that video quality.

Keywords

Streaming video, mobile device, frame splitting, multichannel, dropping rate, and MOS.

3.1 Introduction

Real time video communication over wireless networks faces several challenges such as high error rate, bandwidth variations and limitation, and capability constraints on the handhold devices. Among these, the unreliable and error nature of the wireless channel is the major challenge to stream video over wireless channels [6].

In the case of bad signals ratio, and with high error rates, in the mobile network, the quality of the transmitted video will be affected negatively and the viewer perceives a frozen frame for a certain duration followed by a more or less abrupt change in the picture content due to frame dropping [13, and 12]. It's very hard to guarantee the transmission for all the video frames over single channels, thus, multi channels are proposed by several researchers to increase connection reliability [4] and to enhance the network capacity [5].

Our proposed technique is to overcome the freezing frames in the mobile device and provides a smooth video playing over wireless network. This is done by splitting each frame into two sub-frames containing half of the picture into each. Then the two sub-frames are streamed through two wireless channels. If there is a missing sub-frame from any stream a reconstruction mechanism will take place in the mobile device at its full frame shape. Our subjective test shows that the proposed techniques could be useful to provide a satisfactory quality to the mobile viewer.

3.2 Background and Related Work

Streaming technology delivers media over a network from the server to the client in real time. Streaming video is the classical technique for achieving smooth playback of video directly over the network without downloading the entire file before playing the video [7, 17, and 19]. Streaming video requires high reliability with a low bounded jitter (i.e. variation of delays) and reasonably high transmission rate [8]. Video streaming requires a steady flow of information and delivery of frames by a deadline; wireless radio networks have difficulties to provide such reliably service [20].

The availability of multiple channels for wireless communication provides an opportunity for performance improvement of video application. The term multichannel refers to wireless technology that can use more than one radio channel.

The use of multiple wireless channels has been advocated as one approach for enhancing network capacity. Some wireless devices achieve this property using multi-radio systems, with each interface communicating on a different physical channel. Other devices have just a single radio transceiver, which is tune able to any of the available channels [3]. The use of multiple paths through the transport network for streaming has been proposed to help overcome the loss and delay problems that afflict streaming media and low latency communication. In addition, it has long been known that multiple paths can improve fault tolerance and link recovery for data delivery, as well as provide larger aggregate bandwidth, load balancing, and faster bulk data downloads [1].

Shenoy and Vin [15], suggested that the video server can partition each video stream into two sub-streams (a low-resolution and a residual component stream) in order to support interactivity. During the interactive mode, only the low-resolution stream is transmitted to the client, this can reduces the amount of data that needs to be retrieved and sent to the client’s mobile. Apostolopoulos [1] uses two different paths to send even and odd frames encoded using Multiple Description Coding (MDC). He also suggests that it can be beneficial to send different amounts of traffic on different paths.

Aziz and Lundberg [2], come up with a mechanism to play the complete video frame sequence in the mobile station over error prone channels to eliminate the video freezing. This can be done by transferring the video frames in gray scale, while the second stage is to create duplicated frames that can be transmitted over two channels in the cellular network. After the two video streams have been received by the mobile client, there is a possibility that frames could be missing or corrupted in any stream, to overcome the missing frame or unreadable frame a switching between video streaming channels will take place to make sure that the video player in the mobile device will play the complete video frame sequence, but the main limitation for their work is that the video is played in gray scale and the overhead are increased to double because of the duplicate frames are transmitted over two channels.

3.3 The Proposed Technique

Mobile video streaming is characterized by low resolutions and low bit rates. The bit rates are limited by the capacity of UMTS radio bearer and restricted processing power of mobile terminals; the commonly used resolutions are Quarter Common Intermediate Format (QCIF, 176×144 pixels) for cell phones [14].

Mobile real time application like video streaming suffers from high loss rates over the wireless network [11], and the effect of that on the mobile users may notice some sudden stop during the playing video, the picture momentarily frozen, followed by a jump from one scene to a totally different one.

The use of multiple channels over a single channel is to overcome the problems of limited bandwidth, and to increase the channel capacity for streaming videos [5]. Multichannel has been proposed recently in mobile cellular network by many researcher like [5, and 18], where multichannel provides information to clients via multiple channels.

Our proposed technique is to avoid the freezing picture in the mobile device. The mobile user requests a connection and starts to stream the video. Each frame in the video sequence will be split into two sub-frames. The authors identify three ways for splitting the frames, the first one is based on rows wised splitting, while the second is based on columns wised splitting and the third is pixels splitting, where each sub-frame contains the odd and the even information based on the above. In this study, we will use pixel splitting frames, to creates two sub-frames out of each frame where one sub-frame contains the even pixels and the another contains the odd pixels as shown in Figure 1, and each sub-frame will be queued in different buffers and it will be ready to stream the sub-frames over two wireless channels but the second stream will be delay for 2 seconds after the first stream as in Figure 2. The reason behind that is to minimize the effects of any dropping frames or interruption to the wireless channel under different network condition to the same sub-frames. Streaming the video based on independent transmission on the two channels will be used in order to achieve such purpose, where the frame sequence $f^{'} = \{\ f_1^{'},\ f_2^{'}, \ldots\ f_{n-1}^{'}\}$ will transmitted over the first stream, while, $f^{''} = \{\ f_1^{''},\ f_2^{''}, \ldots\ f_{n-1}^{''}\}$, will transmitted over the second stream.

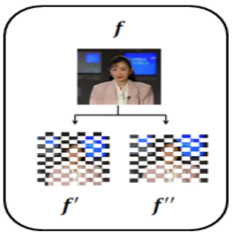


Figure 1: Frame split.

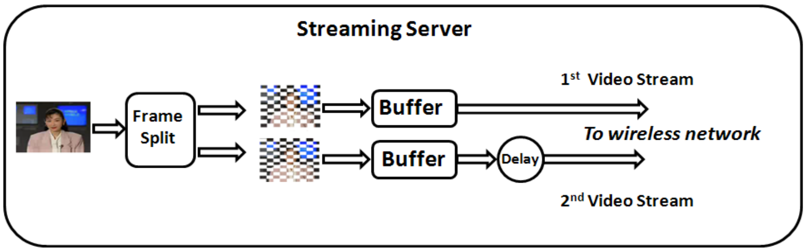


Figure 2: Streaming split sub - frames over two wireless channels.

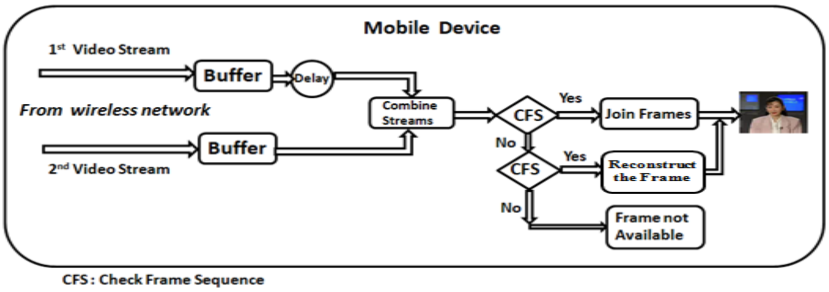


Figure 3: Receiving sub-frames streams of video in the mobile device.

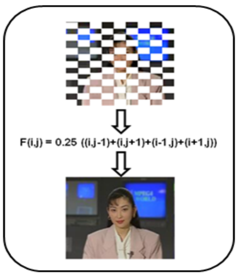


Figure 4: Reconstruct frame.

The mobile station will start receiving the video stream and it will be held in the jitter buffer until the available amounts of frames have been received to start playing. This is for the normal case when there is single channel handling a single stream. According to our proposed system after the first stream has been received by the mobile device it will be hold in the buffer and it will delayed for 2 seconds until the mobile device starts receiving the second stream of video and it will be hold in the other buffer as shown in Figure 3. After both buffers received the right amount of sub-frames the combination of the both stream will take place, as

$$f = \{ (f_1'', f_1'), (f_2'', f_2'), (f_3'', f_3'), \ldots\ldots, (f_{n-1}'', f_{n-1}') \}.$$

After the both sub-frame are combined, a checking procedure will used to check the availability of the sub–frames, as an example, the first Check Frame Sequence (CFS), will check whether the both sub- frames that are related to the same frame are available or not. If both of them are available then join mechanism will applied to return the frame to its original. In case when there is a network interruption, where the sub-frames are corrupted and will be unreadable by the decoder or the sub-frames are dropped. The second CFS will check if there is at least one sub-frame are available or not, if it’s not available then we considered that the frame is dropped from the frame sequence. If there is at least on sub-frame are available, the reconstruction of the sub-frame will take place in the mobile device by taken the average of the neighbouring pixel to replace the missing pixel (this is the reason why we chose pixels splitting), to get fully frame shape, as shown in Figure 4.

3.4 Subjective Viewing Test

3.4.1 Testing Materials and Environments

The video test sequences used in this work were the samples of video sequences Akiyo, Foreman, News, and Waterfalls. The video sequences were chosen because of their deferent characteristics. Each video are coded as 25 frames/second with a resolution of 176 x 144, the transmission rate are 30 frames/second, and the number of frames are transmitted are 1800 frames. The video sequences are shown on 17 FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of 1680 x 1050 pixels. The video sequences for the original and our proposed scenario are displayed with resolution of 176 x 144 pixels in the centre of the screen with black background with duration of 60 seconds for each video sequence.

The Simulink [16], is used to simulate the proposed technique and for three different dropping rates for the same network traffic condition. Under the light traffic and the dropped rate is between 3-4%, for the medium traffic load the dropped frame rate is between 6-7%, and for the high traffic load the dropped frame rate is between 8-9%.

3.4.2 Testing Methods

Following the guidelines outlined in BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R) [9], a subjective experiment has been conducted at Blekinge Institute of Technology in Sweden. Where the user observed two scenarios, the first scenario, when the observer evaluate the normal video transmission over wireless networks with the effect of three different load, and the second scenario, our proposed scenario.

The participated of thirty non-expert viewer in the test subjects were 26 males and 4 females. They were all university staff and students, and their ages range of 20 to 33 of age.

It is well known that peak signal-to-noise ratio (PSNR) does not always rank quality of an image or video sequence in the same way that a human being. There are many other factors considered by the human visual system and the brain [10]. The mean opinion score (MOS) measurements are used to evaluate the video quality.

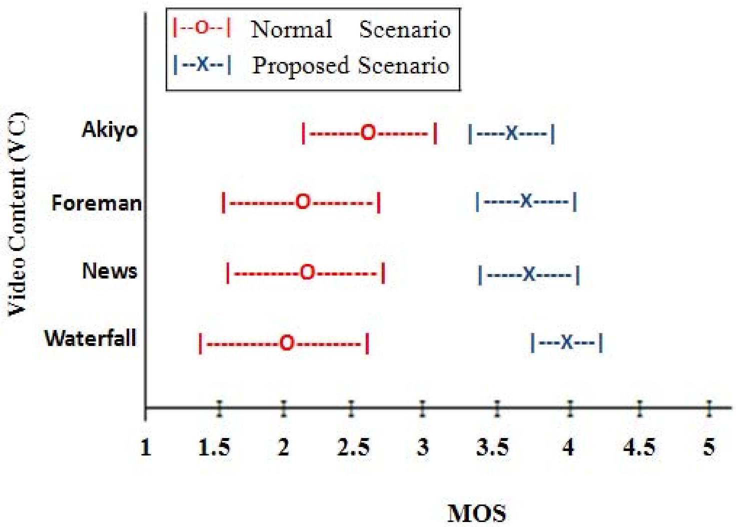
Staff and students evaluated the video quality after each sequence using a five grade MOS scale (1-bad, 2-poor, 3-fair, 4-good, 5-excellent) in a prepared form.

The amount of data gathered from the subjective experiments with respect to the opinion scores that were given by the individual viewers. A concise representation of this data can be achieved by calculating conventional statistics such as the mean score and confidence interval, of the related distribution of scores [9].

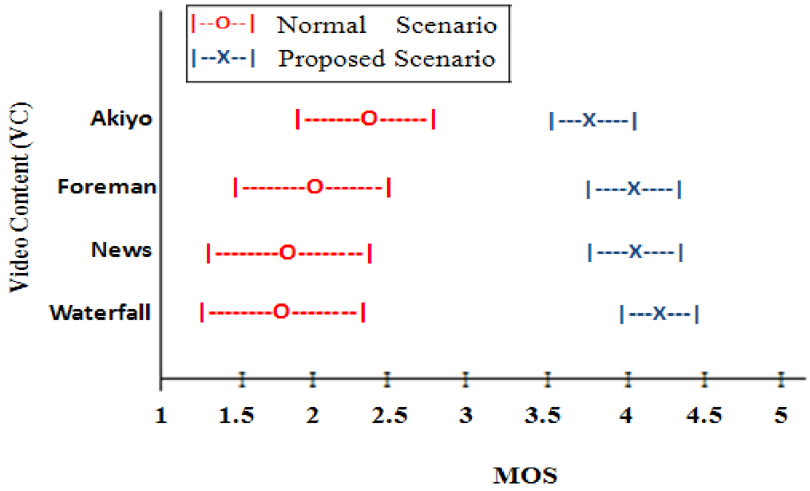
3.5 Experiment Test

The quality of video is subjected to the personal opinion; this means that the quality of service improvements for video transmission has the only goal to satisfy the average human watching the contents of the video. The MOS is obtained through human evaluation tests, where 30 of staff and students are observed the two scenarios. In figure 5, shows the comparison test for the video content (VC) and for different dropping rate percentage, where the centre and span of each horizontal bar indicate the mean score and the 95% confidence interval, respectively.

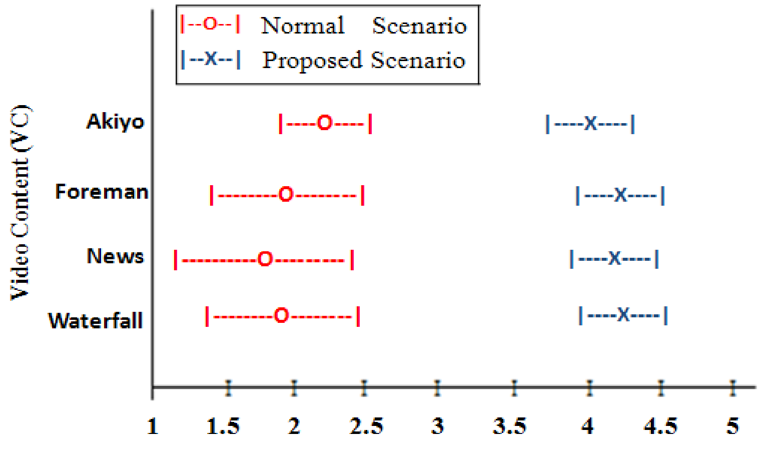
For the normal scenario it can been shown clearly that the observer manage to identify the dropping frames and the frozen picture, where the MOS is lower than 4 corresponding to the five –level quality scale ranks for the light dropping rate and lower than 3 for the medium and the high dropping rate, due to the higher percentage of the dropping frames, which the viewer easily notice the frozen picture.



a. Light dropped rate.



b. Medium dropped rate.



c. High dropped rate.

Figure 5: The MOS for different video contents with different dropping rate.

While for our proposed scenario, the MOS is larger than 3, corresponding to the quality scale ranks and for the three dropping rates. It can be observed that the video present to the viewers resulted in a wide range of perceptual quality ratings indeed for both experiments.

After we analysis their score we feel that our proposed scenario could be a satisfactory technique to eliminate the freezing frames when streaming videos over unreliable network.

3.6 Conclusion

Transmitting a real time video stream over a single channel cannot guarantee that all the frames could be received by the mobile devices. The characteristics of a wireless network in terms of the available bandwidth, frame delay and frame losses cannot be known in advanced. Using multiple channels instead of a single channel is to overcome the problems of limited bandwidth and fading and to increase the channel capacity for streaming videos.

In this work we proposed a frame splitting and streaming technique over two channels under different loads to estimate the effects on a video frame sequence. Our analysis is based on the human opinion and it showed that there is a significant performance improvement for video smoothness under different dropping load over wireless network as compared to traditional techniques. We conclude that our proposed technique appears to provide a promising direction for eliminating the freezing picture for real time transmission under high loss rate and low network capacity channels.

Acknowledgments

We would like to thanks the staff and students of the School of Computing at Blekinge Institute of Technology, Ronneby, Sweden for participating in the subjective experiments and also we would like to thanks the Swedish Knowledge Foundation for sponsoring this work.

References

[1] Apostolopoulos, J. G. 2001. Reliable Video Communication over lose packet networks using multiple state encoding and path diversity. In Proceedings of the Visual Communications and Image Processing. (San Jose, CA, January 24-26, 2001). 392-409.

[2] Aziz, H.M. and Lundberg, L. 2009. Graceful degradation of mobile video quality over wireless network. In Proceedings of the IADIS International Conference Informatics. (Algarve, Portugal, June 17- 19, 2009). WAC'09, 175-180.

[3] Cao, J. and Williamson, C. 2006. Towards stadium - scale wireless media streaming. In Proceedings of the 14th IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems. (Monterey, California September 11-14, 2006). MASCOTS´06, 33-42.

[4] Chen, C. M., Chen, Y. C., and Lin, C. W. 2005. Seamless roaming in wireless networks for video streaming. In Proceedings of the IEEE International Symposium on Circuits and Systems. (Kobe, Japan, May 23-26, 2005). ISCAS ´05, 3255-3258.

[5] Chereddi C., Kyasanur P., Vaidya N. H. 2006. Design and implementation of a multi-channel multi-interface network. In Proceedings of the 2nd International Workshop on Multi-Hop ad Hoc Networks: from Theory to Reality. (Florence, Italy, May 26-26, 2006). REALMAN ´06, 23 – 30.

[6] Dihong T., Xiaohuan L., Al-Regib, G., Altunbasak Y. and Joel R. J. 2004. Optimal packet scheduling for wireless video streaming with error-prone feedback. In Proceedings of the IEEE Wireless Communications and Networking Conference (Atlanta, GA, 21-25March, 2004), WCNC ´04, 1287-1292.

[7] Guangwei B. and Carey W. 2004. The effects of mobility on wireless media streaming performance. In Proceedings of Wireless Networks and Emerging Technologies. (Banff, AB, Canada, July 8-10, 2004). WNET `04, 596-601.

[8] Hsu, C.Y., Ortega, A., and Khansari, M. 1999. Rate control for robust video transmission over burst-error wireless channels. IEEE Journal on Selected Areas in Communications. 17:5,756 – 773.

[9] International Telecommunication Union. Methodology for the subjective assessment of the quality of television pictures. ITU-R, Rec. BT.500-11, 2002.

[10] Martinez-Rach, M., López, O., Piñol, P., Malumbres, M.P., Oliver J., and Calafate, C. T. 2007. Quality assessment metrics vs. PSNR under packet loss scenarios in MANET wireless networks. In Proceedings of the International Workshop on Mobile Video. (Augsburg, Germany, September 24-29, 2007). MV 07, 31-36.

[11] Nguyen, T., Mehra, P., and Zakhor, A. 2003. Path Diversity and Bandwidth Allocation for Multimedia Streaming. In Proceedings of the International Conference on Multimedia and Expo, (Baltimore, Maryland, July 2003). ICME `03, 1-4.

[12] Ong, E. P., We, S., Loke, M. H., Rahardja, S., Tay J., Tan, C. K., Huang, L. 2009. Video quality monitoring of streamed videos. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, (Taipei, Taiwan, April 19-24, 2009). ICASSP `09, 1153 – 1156.

[13] Quan H.-T., M. Ghanbari. 2008. Asymmetrical temporal masking near video scene change. In Proceedings of the IEEE International Conference on Image Processing. (California, San Diego, 12-15 October, 2008). ICIP´08, 2568-2571.

[14] Ries, M., Nemethova O., and Rupp, M. 2007. Performance evaluation of mobile video quality estimators. In Proceedings of the European Signal Processing Conference, (Poznan, Poland, September 3-7 2007). EUSIPCO`07, 159-163.

[15] Shenoy, P. J., and Vin, H. M. 1995. Efficient support for scan operations in video servers. In Proceedings of the 3rd ACM Intl. Conference on Multimedia. (San Francisco, CA, November, 1995). ACM Multimedia'95, 131-140.

[16] DOI=Available: www.mathworks.com.

[17] Xiaozhen C., Guangwei B., and Carey W. 2005. Media streaming performance in a portable wireless classroom network. In Proceedings of IASTED European Workshop on Internet Multimedia Systems and Applications. (Grindelwald, Switzerland, February 21-23, 2005), EuroIMSA ´05, 246-252.

[18] Zheng, B., Wu, X., Jin, X., and Lee, D. L. 2005. TOSA: A near-optimal scheduling algorithm for multi-channel data broadcast. In Proceedings of

the 6th International Conference on Mobile Data Management. (Ayia Napa, Cyprus, May 9-13, 2005). MDM ´05, 29-37.

[19] Zhu, H., Wang, I. H., and Chen, B. 2003. Bandwidth scalable source-channel coding for streaming video over wireless access networks. In Proceedings of Wireless Networking Symposium. (Austin, Texas, October 26-28, 2003). WNCG` 03.

[20] Zhu X., and Girod, B. 2007. Video streaming over wireless networks. In Proceedings of European Signal Processing Conference, (Poznan, Poland, September 3-7, 2007). EUSIPCO`07, 1462-1466.

CHAPTER FOUR

Streaming Video as Space – Divided Sub-Frames over Wireless Networks

Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, and Lars Lundberg

Abstract

Real time video streaming suffers from lost, delayed, and corrupted frames due to the transmission over error prone channels. As an effect of that, the user may notice a frozen picture in their screen. In this work, we propose a technique to eliminate the frozen video and provide a satisfactory quality to the mobile viewer by splitting the video frames into sub-frames. The multiple descriptions coding (MDC) is used to generate multiple bitstreams based on frame splitting and transmitted over multichannels. We evaluate our approach by using mean opinion score (MOS) measurements. MOS is used to evaluate our scenarios where the users observe three levels of frame losses for real time video streaming. The results show that our technique significantly improves the video smoothness on the mobile device in the presence of frame losses during the transmission.

Keywords

Streaming video, mobile device, frame splitting, MIMO, dropping rate, and MOS.

4.1 Introduction

Videos are no longer limited to television or personal computers due to the technological progress in the last decades. Nowadays, many different devices such as laptop computers, PDAs, notebooks and mobile phones support the playback of streaming videos [1].

Streaming video is a technique for smooth playback of video directly over a network without downloading the entire file before playing the video [2, 3, and 4]. Streaming video requires high reliability with a low bounded jitter (i.e. variation of delays) and a reasonably high transmission rate [5]. However, wireless network transmission introduces errors in forms of packet delay, packet inter-arrival time variations, and packet loss, which can a have major impact on the end user experience. This is particularly true in a cellular network environment where the channel condition can vary dramatically [6] and is difficult to estimate [7].

The multiple descriptions coding (MDC) is a source coding method that can generate multiple encoded bitstreams that are equally important and independent. MDC transmits the original video content via different parallel channels. The MDC of a source consists of generating a number of bitstreams (2 or more) that, together, carry the input frames [8, 9, and 10]. The objective of MDC is that if all bitstreams have been received correctly, a high signal quality can be reconstructed, whereas, if some bitstreams have been lost, a low-quality, but acceptable signal quality can still be reconstructed from the received description [11].

Multiple antennae systems with multiple transmitters and multiple receivers, called multiple-input and multiple-output (MIMO) architectures, have been shown to be an effective way to transmit high data rates over wireless channels [12]. MIMO can be used to transmit the video content over multiple wireless channels. In this case, each path may have lower bandwidth, but the total available bandwidths are higher than the single channel. Multi-channel transport can also improve the transport reliability by overcoming the instantaneous congestion problem often encountered in the single-path case [8].

In this paper, we propose a technique to overcome the freezing frame problem in the mobile device and provide a smooth video playback over a wireless network. This is done by streaming the video frames as sub-frames over MIMO architecture, by using the MDC technique and the H.264/AVC codec. If there is a missing sub-frame from any subsequence during the transmission a reconstruction mechanism will be applied in the mobile device to recreate the missing sub-frames and return it to its full frame shape. Our subjective test shows that the proposed technique could be useful to provide smooth playback of the video with a satisfactory quality to the mobile viewer.

4.2 Background and Related Work

Video network traffic is expected to be one of the most important traffic types that need to be supported by high data rates. Video traffic is very hard to manage because it has strict delay and loss requirements.

The hierarchical structure of MPEG streams with possible error propagation through the MPEG frame makes it difficult to send MPEG streams [13]. Some of the received data may become useless to the decoder as insufficient MPEG frame data are available for decoding the MPEG frame when an MPEG frame is dropped [14]. The availability of multiple channels for wireless communication provides an excellent opportunity for performance improvement. The term multichannel refers to wireless technology that can use more than one radio channel. The use of multiple wireless channels has been advocated as one approach for enhancing network capacity.

Apostolopoulos [15], proposed a multiple state video coding, which is designed to combat the error propagation problem that afflicts motion-compensated prediction based coders when there are losses. His approach uses two different paths to send even and odd frames encoded by using MDC. He suggests that it can be beneficial to send different amounts of traffic on different paths. If one stream is lost the other stream can still be decoded to produce usable video. Furthermore, the correctly received streams provide bidirectional (previous and future) information that enables improved state recovery for the corrupted stream.

Zheng et al. [12] proposed a scheme that integrates MDC, hybrid space-time coding structure for robust video transmission over MIMO-OFDM system. The MDC will generate multiple bitstreams of equal importance which are very suitable for multiple-antennas system. They considered a MIMO system with 4 transmitters and 4 receiver antennas for robust video transmission, thus transmitting signals for different subcarriers simultaneously over all transmit antennas. Data partition is used to divide the encoder signal bitstream into two components.

In this scheme, each description is further divided into two partitions: motion vectors and Group of Picture (GOP) header information, and texture information. By transmitting the most critical information using a MIMO diversity scheme, the average reconstructed quality at the receiver will not significantly degrade.

In our previous work [16], we proposed a mechanism to split the video frame into two sub-frames and streaming it over two channels by using Simulink. In this paper our approach is to split the video frame into four sub-frames based on the use of MDC with a compatible H.264/AVC codec [17] and transmitted over 4x4 MIMO architecture.

4.3 The Proposed Technique

Mobile real time applications like video streaming suffer from high loss rates over wireless networks [10]. The result is that the users may notice a sudden stop during the video playing. The picture is momentarily frozen, followed by a jump from one scene to a totally different one.

Our proposed technique is to split each video frame to four sub-frames based on a pixel distribution according to Figure 1 and 2, respectively, where each sub-frame contains one fourth of the main frame pixels. The four sub-frames will be encoded by MDC using a H.264/AVC codec. The encoded sub-frames will be transmitted over a MIMO architecture.

4.3.1 Encoding the Sub-Frames

The input video frames are split into four sub-frames, where each sub-frame will represent its own subsequence. The first subsequence is transmitted without any delay, the second subsequence will be delayed for 0.5 seconds; the third subsequence will be delayed for 1.0 seconds, while the fourth subsequence will delayed for 1.5 seconds as shown in Figure 3. The reason for implementing the subsequence transmission delay is to minimize the effect of any dropping or corruption to the sub-frames that belong to the same frame over a wireless channels and under different network condition.

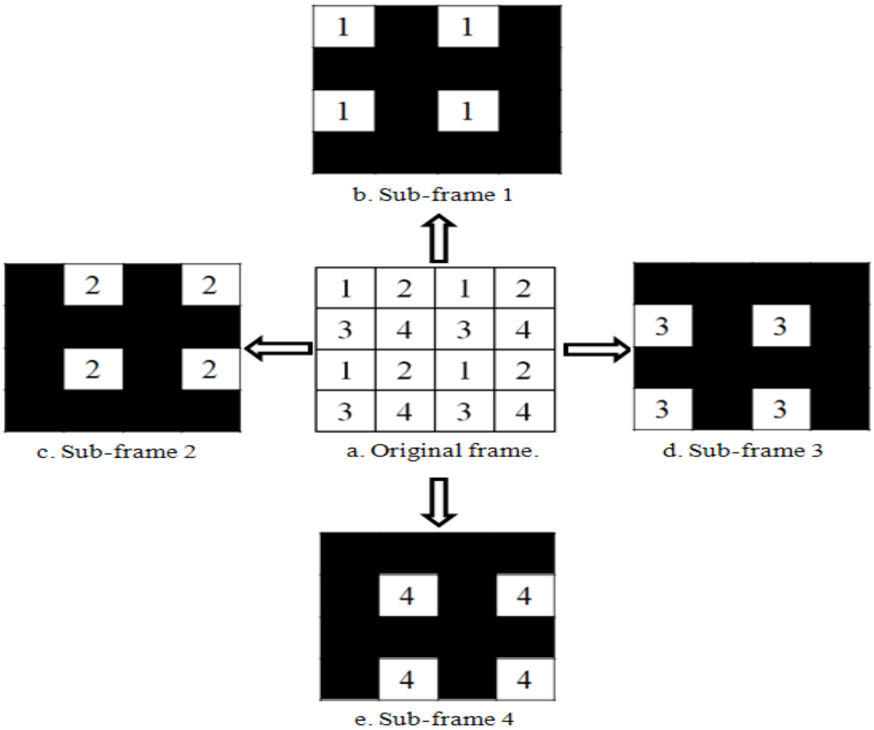


Figure 1: Frame splitting.

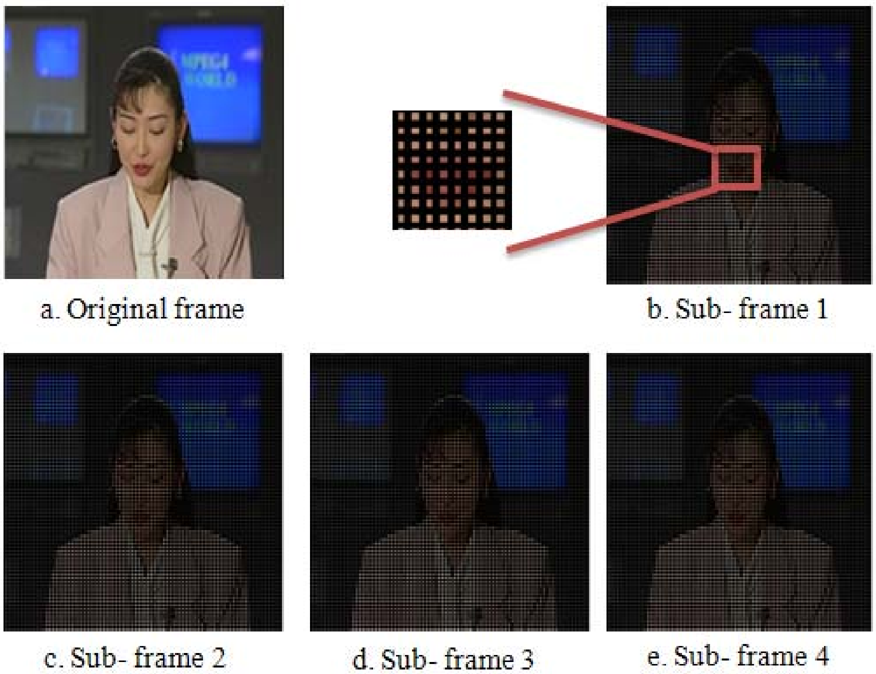


Figure 2: Snapshot of Akiyo frame splitting.

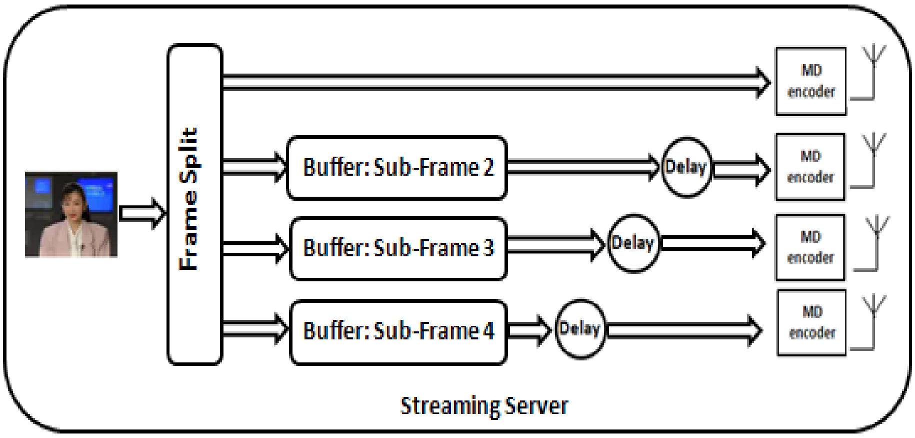


Figure 3: Streaming the sub-frames over multichannels.

4.3.2 Decoding the Sub-Frames

In the normal case, when the streaming video is transmitted over a single channel, the mobile device will start receiving the video frames and it will be held in the buffers until the right number of frames has arrived to start playing the video.

In our proposed technique, after the first subsequence has been received by the mobile device it will be held in the buffer and it will be delayed for 1.5 seconds, while the second subsequence will be held in another buffer and it will be delayed for 1.0 seconds. The third subsequence will be held in a third buffer and it will be delayed for 0.5 seconds. After the fourth subsequence has been received, the check frame sequence (CFS) procedures will take place, to check the availability of the sub–frames after grouping the sub-frames that are related to the same original frame, as shown in Figure 4.

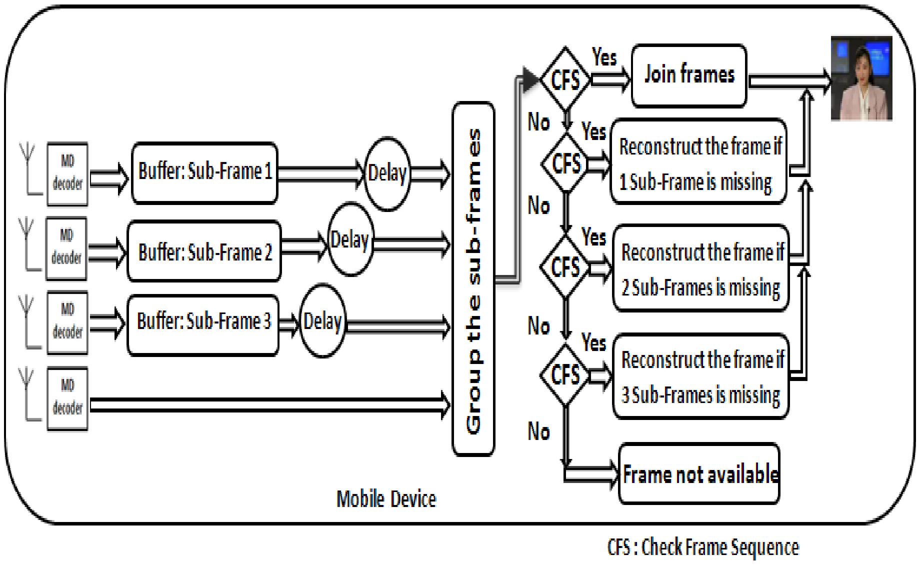


Figure 4: Receiving the sub-frames of the video in the mobile device.

The CFS and the reconstruction mechanism are used to identify the missing sub-frames and build the frame to its normal shape. This is done according to the following checking procedures:

- The first CFS will check whether all the sub-frames that are related to the same original frame are available. If the four related sub-frames are available then a joining mechanism will be applied to return the frame to its original shape.
- The second CFS will check if there are at least 3 sub-frames are available as shown in Figure 5. If one sub-frame is missing then the average of the neighbouring pixels will be calculated to replace the missing frame pixels and return the full frame to its normal shape.
- The third CFS will check if there are at least 2 sub-frames available as shown in Figure 6. If two sub-frames are missing then the average of the neighbouring pixel will be calculated to replace the missing sub-frame pixels and return the full frame to its normal shape.

- The fourth CFS will check if there is at least 1 sub-frame is available as shown in Figure 1. If three sub-frames are missing then the average of the neighbouring pixel will be calculated twice, the first time to find the half of the frame as shown in Figure 6 and the second time to return the full frame to its normal shape.

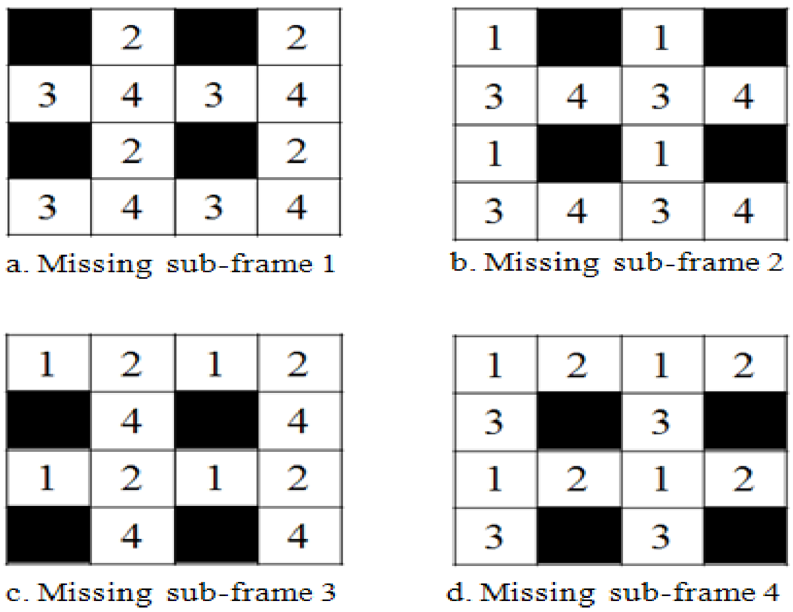


Figure 5: The possibility of missing one sub-frame.

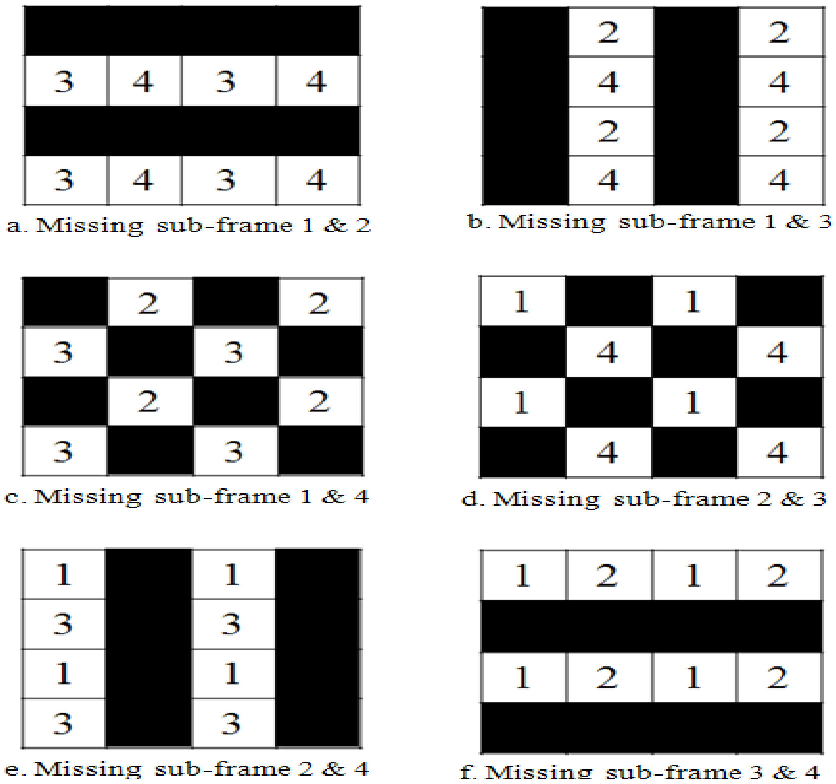


Figure 6: The possibility of missing two sub-frames.

4.4 Subjective Viewing Test

4.4.1 Testing Methods

It is well known that peak signal-to-noise ratio (PSNR) does not always rank quality of an image or video sequence in the same way as a human being. There are many other factors considered by the human visual system and the brain [18]. One of the most reliable ways of assessing the quality of a video is subjective evaluation using the mean opinion score (MOS). MOS is a subjective quality metric obtained from a panel of human observers. It has been regarded for many years as the most reliable form of quality measurement technique [19].

The MOS measurements that are used to evaluate the video quality in this study follow the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). The physical Lab, with controlled lighting and set-up, conforms to the ITU-R recommendation. The score grades in this method range from 0 to 100 which is mapped to the quality ratings on the 5-grade discrete category scale labeled with Excellent (5), Good (4), Fair (3), Poor (2) and Bad (1) [20].

The data gathered from the subjective experiments reflect the opinion scores that were given by the individual viewers. A concise representation of this data can be achieved by calculating conventional statistics, such as the mean score and confidence interval, of the related distribution of scores. The statistical analysis of the data from the subjective experiments reflects the fact that perceived quality is a subjective measure and hence will be described statistically according to the ITU-R guidelines [20].

4.4.2 Testing Materials and Environments

The simulation study for the proposed technique is based on the combination of the MDC/MIMO transmission schemes, using the H.264 ffmpeg codec [17] for the video test sequences Akiyo, Foreman, News, and Waterfall [21]. The video sequences were chosen because of their characteristics. Each video is encoded as 25 frames/second with a resolution of 176 x 144, the transmission rate is 30 frames/second, and the total number of frames transmitted is 1800.

We evaluate our system using different drop rates, i.e., the fraction of the transmitted frames that are lost during the transmission. Under light traffic the drop rate is 3%, and the length duration for the frame loss is 20 frames. For medium traffic load the drop rate is 6%, and the length duration for the frame loss is 40 frames. While for high traffic load the drop rate is 9%, and the length duration for the frame loss is 60 frames.

The video sequences are shown on a 17 inch EIZO FlexScan S2201W LCD computer display monitor with a native resolution of 1680 x 1050 pixels. The video sequences for the frozen and our proposed scenarios are displayed with resolution of 176 x 144 pixels in the centre of the screen with a black background with a duration of 60 seconds for each video sequence.

4.5 Experiment Results

The experienced quality of video is subject to the personal opinion; where the quality of service (QoS) improvements for video transmission has the only goal to satisfy an average human watching the contents of the video stream.

The subjective evaluation was conducted at Blekinge Institute of Technology in Sweden. We used thirty non-expert test subjects, 27 males and 3 females. They were all university staff and students and with an age range of 22 to 33. The MOS is obtained through human evaluation tests, using three different scenarios with three different frame drop rates:

- The first scenario, the observers evaluate the video stream over a single wireless channel using the frozen picture technique to stream the video and based on the three frame dropping rates.
- The second scenario, the observers evaluate the video stream over multichannels, using our proposed technique, where one sub-frame is missing and then reconstructed.
- The third scenario, the observers evaluate the video stream over multichannels using our technique where three sub-frames are missing and then reconstructed.

The results of the scenario where two sub-frames are missing from the original frame are not included as the results have been reported by [16]. The snapshot for the reconstructions of the missing sub-frames based on the above scenarios and the scenario in [16] to the videos frames, Akiyo, Foreman, News, and Waterfall are shown in Figure 7, 8, 9, and 10 respectively. Figure 11 and 12; show the comparison test for the video content (VC) and for different dropping rates percentage, where the centre and span of each horizontal bar indicate the mean score and the 95% confidence interval, respectively.

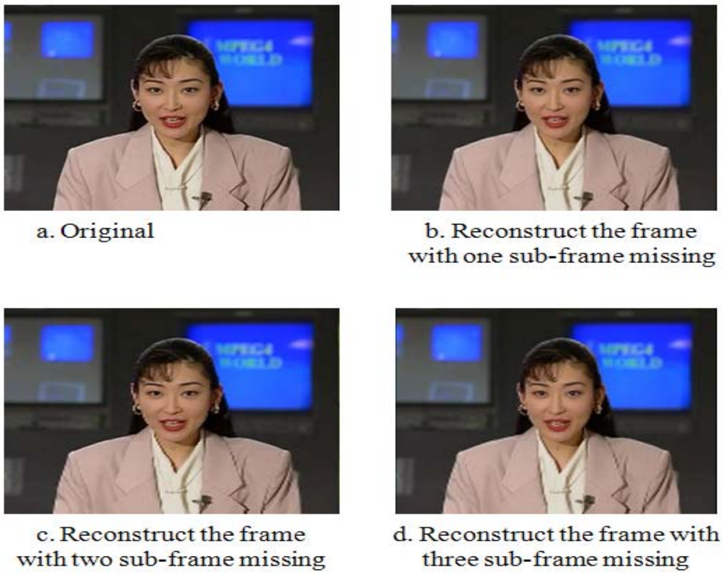


Figure 7: Snapshot of Akiyo video frame number 140.

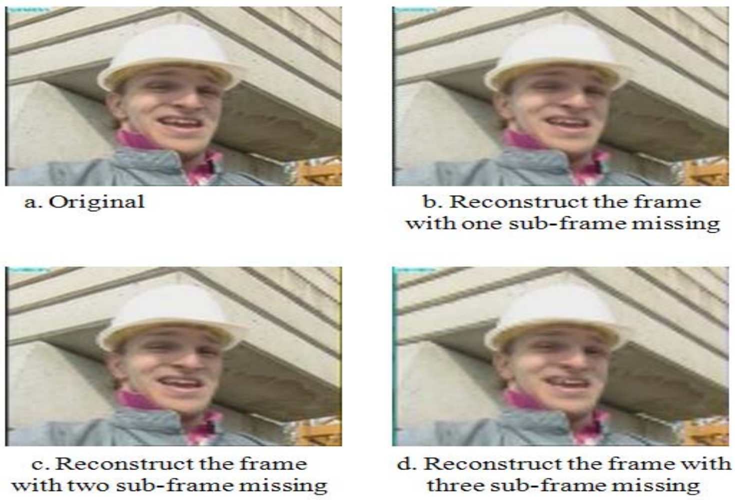
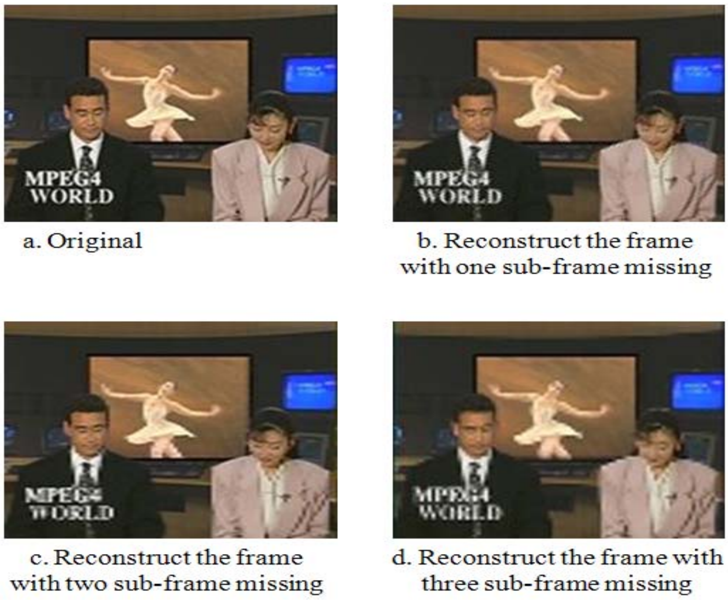


Figure 8: Snapshot of Foreman video frame number 140.



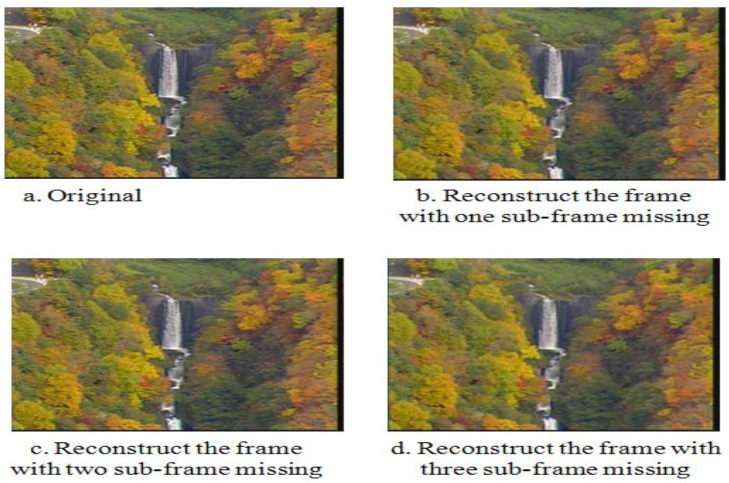
a. Original

b. Reconstruct the frame with one sub-frame missing

c. Reconstruct the frame with two sub-frame missing

d. Reconstruct the frame with three sub-frame missing

Figure 9: Snapshot of News video frame number 140.



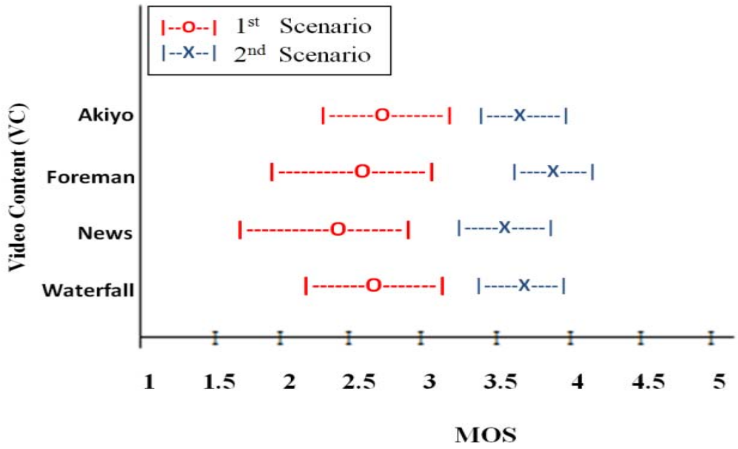
a. Original

b. Reconstruct the frame with one sub-frame missing

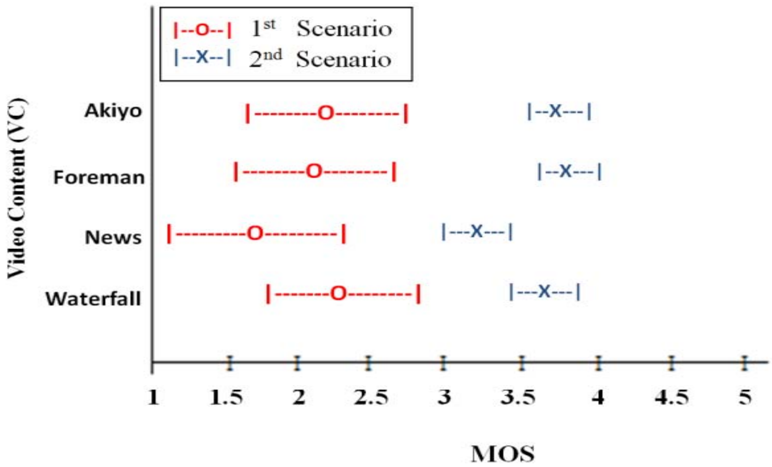
c. Reconstruct the frame with two sub-frame missing

d. Reconstruct the frame with three sub-frame missing

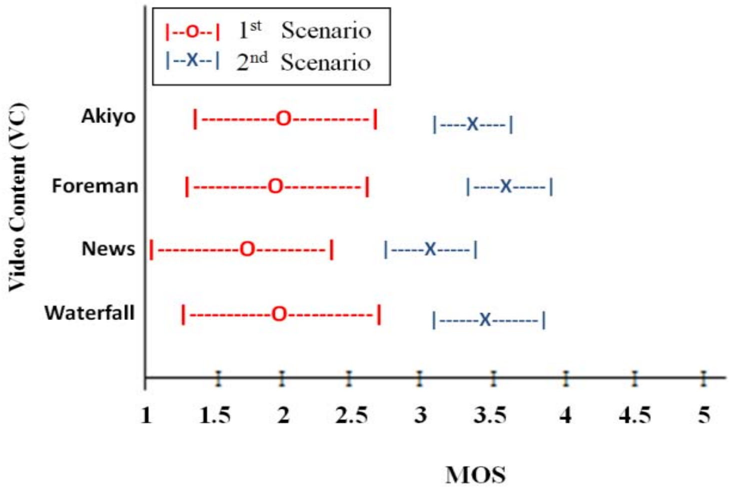
Figure 10: Snapshot of Waterfall video frame number 140.



a. Light drop rate.

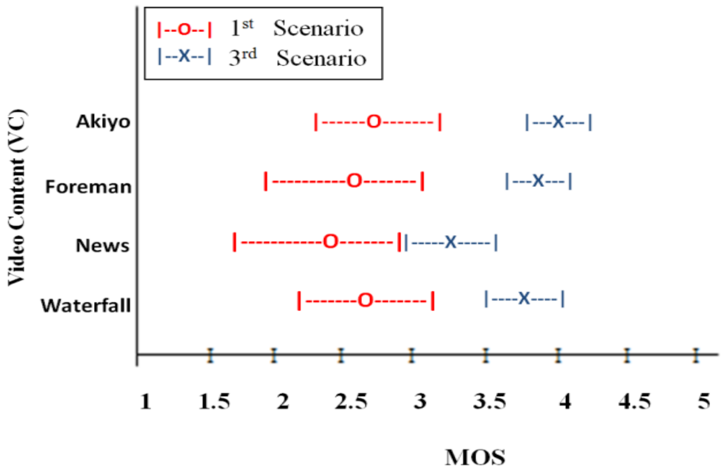


b. Medium drop rate.

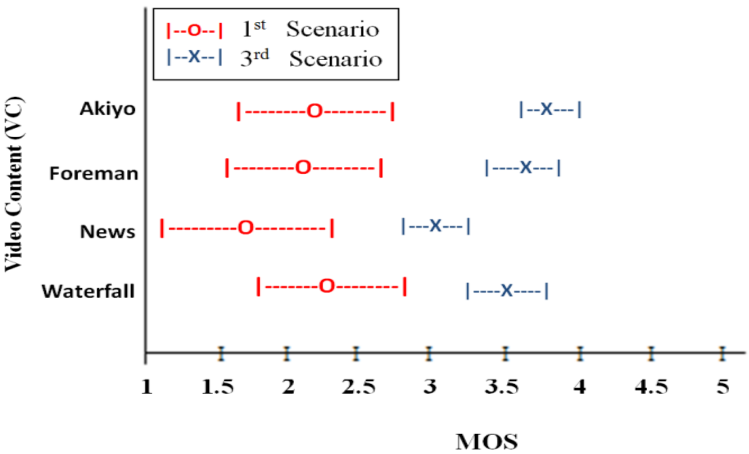


c. High drop rate.

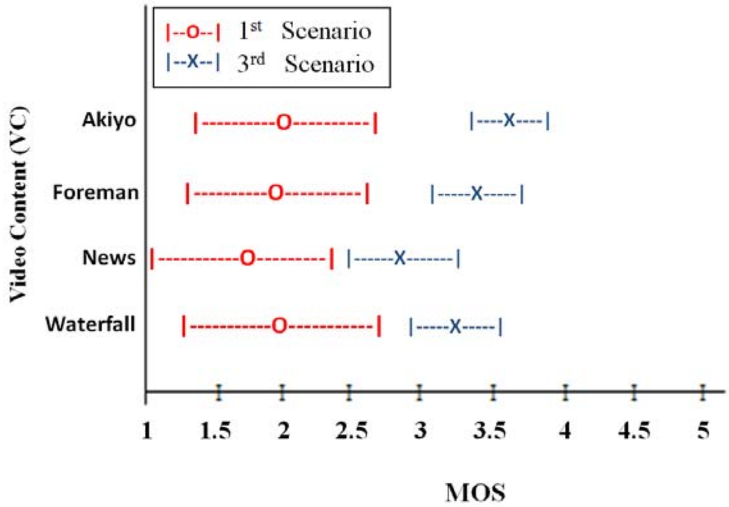
Figure 11: The MOS comparison between scenario 1 and 2 and for different video contents under different frame dropping rates.



a. Light drop rate.



b. Medium drop rate.



c. High drop rate.

Figure 12: The MOS comparison between scenario 1 and 3 and for different video contents under different frame dropping rates.

In the first scenario it can be clearly seen that the observer manages to identify the dropping frames and the frozen picture, where the MOS is lower than 3.5 for the light dropping rate based on the five–level quality scale ranks. For medium and high dropping rate the MOS is lower than 3, due to the higher percentage of the frame dropping rate, in which the viewer easily notices the frozen picture.

In the second scenario, the MOS is larger than 3 for the light and medium dropping rate percentage. While for the high dropping rate percentage the MOS is larger than 3 except the News streaming video is larger than 2.5 as shown in Figure 11, due to the effect of percentage of the sub-frame lost and the reconstruction mechanism as shown in Figure 9(b).

In the third scenario, the MOS is larger than 3 for the light, medium and high dropping rate percentage, except the News streaming video is larger than 2.5 as shown in Figure 12, due to the effect of percentage of the sub-frame lost and the reconstruction mechanism as shown in Figure 9(d).

It can be observed that the video presented to the viewers resulted in a wide range of perceptual quality ratings for both experiments, as shown in Figure 11 and 12 respectively. In general, we observe that our proposed technique in all cases have a higher MOS that the frozen picture technique. Therefore, we conclude that our proposed technique is a satisfactory technique to eliminate the freezing frames when streaming videos over unreliable network.

4.6 Conclusion

In this work we proposed a technique to address the frozen picture problem when streaming videos over mobile network. A frame splitting mechanism splits each video frame into four sub-frames. The sub-frames then streamed over MIMO architecture. The sub-frames are joined together in the mobile device, and a reconstruction mechanism is applied to the available sub-frames to return the frame to its normal shape; when there are missing sub-frames.

Our evaluation is based on the human opinion using subjective evaluations based on the mean opinion score (MOS). The results show that there is a significant performance improvement for video smoothness under different frame drop rates over a wireless network as compared to the traditional techniques.

We conclude that our proposed technique appears to provide a promising direction for eliminating the freezing picture problem for the mobile device viewers and for real time transmission under high frame loss rates. However, the quality of the receiving video is degraded.

Acknowledgments

We would like to thanks the staff and students at Blekinge Institute of Technology, Ronneby, Sweden for participating in the subjective experiments and also we would like to thank the Swedish Knowledge Foundation for sponsoring a part of this work through the project QoEMoS (21601420).

References

[1] S. Kopf, T. King, F. Lampi, and W. Effelsberg, “Video color adaptation for mobile devices,” In Proc. of the 14th ACM Int’l Conf. on Multimedia, pp. 963-964, Oct., 2006.

[2] B. G. Wei, and W. Carey, “ The effects of mobility on wireless media streaming performance,” In Proc. of the Wireless Networks and Emerging Technologies, WNET `04, pp. 596-601, July, 2004.

[3] C. Xiaozhen, B. Guangwei, and W. Carey, “ Media streaming performance in a portable wireless classroom network,” In Proc. of the IASTED European Workshop on Internet Multimedia Systems and Applications, Euro IMSA ´05, pp. 246-252, Feb., 2005.

[4] Z. Hua, G. Zeng, I. Chlamtac, “ Bandwidth scalable source-channel coding for streaming video over wireless access networks,” In Proc. of the Wireless Networking Symposium, WNCG` 03, Oct., 2003.

[5] C. Y. Hsu, A. Ortega, and M. Khansari, “Rate control for robust video transmission over burst-error wireless channels,” IEEE Journal on Selected Areas in Communications. 17(5): 756 - 773, 1999.

[6] Weber, M. Guerra, S. Sawhney, L. Golovanvsky, and M. Kang, “Measurement and analysis of video streaming performance in live UMTS networks,” In Proc. of the Int’l Symp. Wireless Personal Multimedia Communications, WPMC‘06, pp: 1-5, Sept., 2006.

[7] J. Li, and L. Li, “Research of transmission and control of real-time MPEG-4 video streaming for multi-channel over wireless QoS mechanism,” In Proc. of the First Int’l Multi-Symp. of Computer and Computational Sciences, IMSCCS ’06, vol. 2, pp. 257-261, June, 2006.

[8] J. Zhou, H-R Shao, C. Shen, M-T Sun, “Multi-path transport of FGS video, packet video (PV),” April 2003. Technical report. http://www.merl.com.

[9] W. Dapeng, Y. T. Hou, W. Zhu, Y-Q Zhang, and J. M. Peha, “Streaming video over the internet: approaches and directions,” IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Streaming Video, 11(3): 282-300, 2001.

[10] Y. Wang, A. R. Reibman, S. Lin, “Multiple description coding for video delivery”, In Proc. of the IEEE Journal , 93(1): 57-70, 2005.

[11] L. V. Andrea, A. Borneo, F. Marco, R. Roberto, “Video over IP using standard-compatible multiple description coding: an IETF proposal,” Journal of Zhejiang University Science, 7(5): 668-676, 2006.

[12] H. Zheng, C. Ru, C.W. Chen, L. Yu, “Video transmission over MIMO-OFDM System: MDC and space-time coding-based approaches,” Hindawi Publishing Corporation, Advances in Multimedia, vol. 2007, pp. 1-8, 2007.

[13] A. Ziviani, B. E. Wolfinger, J. F. Rezende, O. C. Duarte, and S. Fdida, “ Joint adoption of QoS schemes for MPEG streams,” Journal of Multimedia Tools and Applications, 26(1): 59–80, 2005.

[14] M. Krunz, “Bandwidth allocation strategies for transporting variable bit-rate video traffic,” IEEE Communications Magazine, 37(1): 40-66, 1999.

[15] J. G. Apostolopoulos, “Reliable video communication over lose packet networks using multiple state encoding and path diversity,” In Proc. of the Visual Communications and Image Processing, pp. 392-409, Jan., 2001.

[16] H. M. Aziz, H. Grahn, and L. Lundberg, “Eliminating the freezing frames for the mobile user over unreliable wireless networks,” In the Proc. of the 6th Int’l Conf. on Mobile Technology, Applications and Systems; (ACM Mobility), Sept., 2009.

[17] http://www.ffmpeg.org.

[18] M. Martinez-Rach, O. López, P. Piñol, M. P. Malumbres, J. Oliver, and C. T. Calafate, “Quality assessment metrics vs. PSNR under packet loss scenarios in MANET wireless networks,” In Proc. of the Int’l Workshop on Mobile Video, MV 07, pp. 31-36, Sept., 2007.

[19] M.Martinez-Rach, O.Lopez, P.Pinol, M.P. Malumbres, J. Oliver, Carlos T. Calafate, “Behavior of quality assessment metrics under packet losses on Wireless Networks,” XIX Jornadas de Paralelismo, Sept., 2008.

[20] International Telecommunication Union, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R, Rec. BT.500-11, 2002.

[21] http://trace.eas.asu.edu/yuv/index.html.

CHAPTER FIVE

Sub-Frame Crossing for Streaming Video over Wireless Networks

Hussein Muzahim Aziz, Håkan Grahn, and Lars Lundberg

Abstract

Transmitting a real time video streaming over a wireless network cannot guarantee that all the frames could be received by the mobile devices. The characteristics of a wireless network in terms of the available bandwidth, frame delay, and frame losses cannot be known in advanced. In this work, we propose a new mechanism for streaming video over a wireless channel. The proposed mechanism prevents freezing frames in the mobile devices. This is done by splitting the video frame in two sub-frames and combines them with another sub-frame from different sequence position in the streaming video. In case of lost or dropped frame, there is still a possibility that another half (sub-frame) will be received by the mobile device. The receiving sub-frames will be reconstructed to its original shape. A rate adaptation mechanism will be also highlight in this work. We show that server can skip up to 50% of the sub-frames and we can still be able to reconstruct the receiving sub-frame and eliminate the freezing picture in the mobile device.

Keywords

Streaming video; wireless network; frame splitting; sub-frame crossing; sub-frame merging; rate adaptation.

5.1 Introduction

Mobile cellular networks provide freedom to the mobile users for calling and receiving calls anywhere and at any time while the mobile users are on the move. Streaming technology delivers video frame over a network from the server to the client in real time. Streaming video is the classical technique for achieving smooth playback of video directly over the network without downloading the entire file before playing the video [1]-[3]. Streaming video requires high reliability with a low bounded jitter (i.e. variation of delays) and reasonably high transmission rate [4].

Real-time video streaming requires a steady flow of information and delivery of frame packets according to their deadline; wireless radio networks have difficulties to provide such reliable service [5]. However, the transmission errors in forms of packet delay, jitter (packet inter-arrival variations), and packet loss can have major impact on the end user experience. This is particularly true in a cellular network environment where the channel condition can vary dramatically due to fading and other network effects [6].

Bandwidth is one of the most critical resources in wireless networks, and thus, the available bandwidth of wireless networks should be managed in an efficient manner [7]. Therefore the transmission rate of streaming video should be reduced according to the bandwidth of the communication networks [8][9]. Network adaptation refers to how many network resources (e.g., bandwidth) a video stream should utilize for video content, resulting in designing an adaptive streaming mechanism for video transmission [10]. To stream video, it is desirable to adjust the transmission rate according to the perceived congestion level in the wireless network to maintain a suitable loss level and fairly shared bandwidth with other connections. Furthermore, it's favourable for the streaming video to be aware of the transmission level in the wireless channel in order to obtain good streaming quality by appropriate error protection.

Aziz et al. [11], present a mechanism to split the video frame into two sub-frames and transmit it over two wireless channels in a cellular network. After the two video streams have been received by the mobile device, a checking procedure will used to identify the missing sub-frames. The checking procedure is needed to identify the missing sub-frame and to reconstruct the available sub-frame. The reconstruct mechanism is based on taking the average of the neighbouring pixels to replace the missing sub-frame to get fully frame shape, but the main limitation for their work is transmitting each video frame over two wireless channels.

To overcome the transmission of each frame over two wireless channels, a sub-frame crossing mechanism is proposed based on frame splitting mechanism in [11]. It will split the frame in two sub-frames, and combine the sub-frame with another sub-frame from a different sequence position and transmitted as a single frame over a single wireless channel. In case of frame losses or frames corruption from the streaming video, there is still a possibility that one of the sub-frame will be received by the mobile device. The receiving sub-frame will be reconstructed to its full frame shape.

5.2 The Proposed Technique

Mobile video streaming is characterized by low resolutions and low bit rates. The bit rates are limited by the capacity of UMTS radio bearer and the restricted processing power of mobile terminals. The commonly used resolutions are Quarter Common Intermediate Format (QCIF, 176x144 pixels) for cell phones [12]. Mobile real time application like video streaming suffers from high loss rates over the wireless network [13] and the effect of that the mobile users may notice some sudden stop during the playing video, the picture is momentarily frozen. Frozen pictures could occur if the video frames have been delayed or been dropped.

The proposed mechanism is based on the way of packetizing the video frame and transmitted over a single wireless channel and this can be done by splitting the video frame into two sub-frames based on the pixel splitting technique [11] to create two sub-frames out of each frame, where one sub-frame contains the even pixels and another contains the odd pixels as shown in Fig. 1[11].

The mobile client's request the connection to the server, the server will start streaming by splitting each frame into two sub-frames which will be queued in different buffers according to the frames sequence number and to its contains.

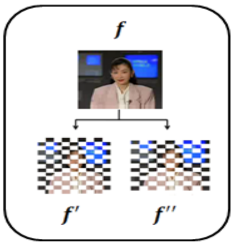


Figure1: Frame split.

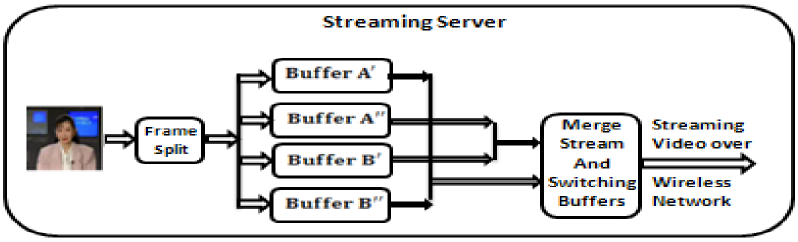


Figure 2: Streaming crossing sub-frames over wireless channels.

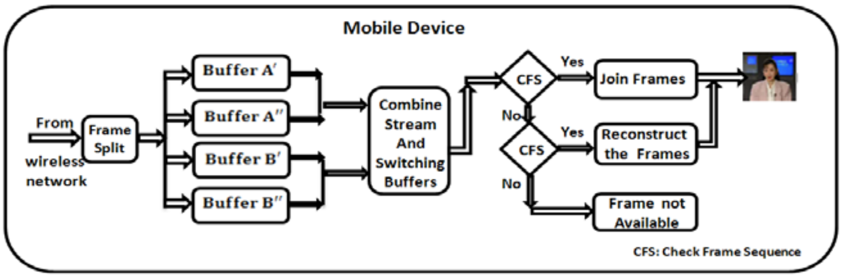


Figure 3: Receiving the streaming sub-frames crossing in the mobile device.

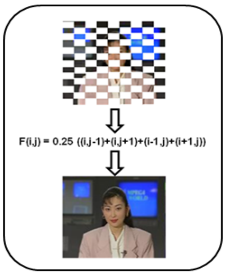


Figure 4: Reconstruct frame.

In this study, we will assume that the transmission rate is 30 frames/second. The first 15 frames ($f_{1}, f_{2}, f_{3}, \ldots, f_{15}$) will be queued in two buffers and each buffer will hold 15 sub-frames as an example Buffer $A' = (f'_{1}, f'_{2}, f'_{3}, \ldots, f'_{15})$, $A'' = (f''_{1}, f''_{2}, f''_{3}, \ldots, f''_{15})$, while the next 15 frames ($f_{16}, f_{17}, f_{18}, \ldots, f_{30}$) will also be queued in two buffers as $B' = (f'_{16}, f'_{17}, f'_{18}, \ldots, f'_{30})$, $B'' = (f''_{16}, f''_{17}, f''_{18}, \ldots, f''_{30})$, as shown in Fig. 2.

The crossing mechanism between two sub-frames from different buffers and from different sequence position will be combined with each other and will create as a normal frame and as an example $\{(f'_{1}, f''_{16}), (f'_{2}, f''_{17}), \ldots \ldots \ldots, (f'_{14}, f''_{29}), (f'_{15}, f''_{30}), (f'_{16}, f''_{1}), \ldots, (f'_{29}, f''_{14}), (f'_{30}, f''_{15})\}$.

This will be applied to the entire video frame and will be transmitted over a single channel. The reason behind that, if there is a lost or dropped frame from the streaming video, the effect will be on two sub-frames from two different frames that are in different positions from the streaming sequence.

After each frame has been received by the mobile device a splitting frame technique will applied. The sub-frames will be hold in different buffers and according to the way they have been split at the server side as shown in Figure 3. The sub-frame will distributed to the relevant buffers and a combination of the sub-frames according to their sequence will be considered with buffers switching to create frame sequences of the stream.

A frame checking sequence will take place after the combination between the sub-frames that are related to the original frame. A checking procedure will used to check the availability of the sub-frames, as an example, the first check frame sequence (CFS), will check whether the both sub- frames that are related to the same frame are available or not. If both of them are available then a join mechanism will be applied to return the frame to its original. In the case when there is a network interruption, e.g., if the sub-frames are corrupted and will be unreadable by the decoder or the sub-frames are dropped, the second CFS will check if there is at least one sub-frame available or not.

If it’s not available then we considered that the frame is dropped from the frame sequence. If there is at least on sub-frame are available, the reconstruction of the sub-frame will take place in the mobile device by taken the average of the neighbouring pixel to replace the missing pixel and to get the normal frame shape and play it on the mobile screen with acceptable quality level, as shown in Figure 4 [11]. While the experimental study for the sub-frame crossing mechanism has been tested according to International

Telecommunication Union (ITU-R) [14] and under different network condition and for different motion video and we get similar results as in [11].

5.3 Rate Adaption

Rate adaptation for streaming video is regarded as a necessary mechanism in order to handle the network conditions, and fluctuations of available network bandwidth. During the interactive mode, the server rate is adapted to the requirement of the available bandwidth in the wireless network.

The adaption rate for the sub-frame crossing mechanism should be considered carefully to avoid skipping the sub-frames that belonging to the same video frame and with the consideration of the network bandwidth and network interruption to the streaming video. This can be done by not considering the combination of the two sub-frame and transmitting only one half of the frame (any sub-frame) to the mobile device and according to the following percentages:

- 10% adjustment: we will skip every 5th sub-frames from buffer A' and B'.
- 25% adjustment: we will skip every even sub-frames from buffer A' and B'.
- 50% adjustment: we will skip all the sub-frames from buffer A' and B'.

The rate adaptation mechanism is needed to adjust the transmission rate based on the congestion level. The server will adjust the transmission rate by skipping the sub-frames that are not related to each other and the skipping rate limits shouldn't cross 50% from streaming video to avoid discarding the sub-frames that are related to the same video frame.

5.4 Discussion and Conclusions

The high error rate and network congestion raises a number of challenging issues for streaming video over unreliable wireless network. Such issues include techniques to ensure a continuous display at the mobile device in spite of errors in the received video. For wireless networks in particular, the available channel bandwidth varies with time that can potentially result in lost frames or frames arriving much later than its deadline, resulting in garbled frame.

Sub-frame crossing mechanism is proposed to distribute the sub-frames in different positions in the streaming sequence by combine it with another sub-frames. The idea behind that is to avoid losing the complete frame and allow at least half of the frame (sub-frame) to be received by the mobile device.

The proposed mechanism is based on the limitation on [11], where the analysis comparison between the two scenarios is presented in Table 1.The proposed mechanism is based on packetizing the video frames and transmitted over a single channel. The extra implement delay time is needed for the sub-frame to merge with another sub-frame from different position in the server with the amount of time is needed in the mobile device to combine the sub-frame to the original frame shape is 1 second. While the discard frames rates should be limited to 50% from the amount of transmission rate been streamed from the server.

Table 1: Implementation comparison

	Proposed	**In [11]**
No. of Channel(s)	1	2
Overhead	Normal	Double
Extra Delay Time Implementation	1 Sec.	2 Sec.

Streaming the sub-frame over tow channels scenario [11] will increase the overhead to double; while the implementation delay time been applied is 2 seconds. Implementing the delay time is to avoid the transmission of the same sub-frame at the same time under the same network condition and this will minimize the effects on the same frame.

While the combination between sub-frames will take place after the mobile devices start receiving the same amount of sub-frames from both channels. While the main strength is that if the amount of skipped frames are high it’s still there is a possibilities that we could reconstruct the missing sub-frame.

Our proposed mechanism appears to provides a promising direction for eliminating the freezing picture in the mobile screen, that been caused by the missing frames from the streaming sequence. Adjusting the amount of frames to be transmitted according to the changes in the bandwidth is highly needed to reduce the amount of frames that needs to be send to the mobile device under congested network. However, the quality of the receiving video is degraded.

References

[1] B. Guangwei, and W. Carey,” The effects of mobility on wireless media streaming performance,” In Proceedings of Wireless Networks and Emerging Technologies. Banff, AB, Canada, WNET `04, pp: 596-601, July 8-10, 2004.

[2] C. Xiaozhen, B. Guangwei, and W. Carey, ”Media streaming performance in a portable wireless classroom network,” In Proceedings of IASTED European Workshop on Internet Multimedia Systems and Applications. EuroIMSA ´05. Grindelwald, Switzerland, pp: 246-252, Feb. 21-23, 2005.

[3] H. Zhu, I. H. Wang, and B. Chen, “Bandwidth scalable source-channel coding for streaming video over wireless access networks,” In Proceedings of Wireless Networking Symposium. WNCG` 03. Austin, Texas, Oct. 26-28, 2003.

[4] C. Y. Hsu, A. Ortega, and M. Khansari, “Rate control for robust video transmission over burst-error wireless channels,” IEEE Journal on Selected Areas in Communications. vol. 17, no. 5, pp: 756 – 773, 1999.

[5] X. Zhu, and B. Girod, “Video streaming over wireless networks,” In Proceedings of European Signal Processing Conference. EUSIPCO`07. Poznan, Poland, pp: 1462-1466, Sept. 3-7, 2007.

[6] R. Weber, M. Guerra, S. Sawhney, L. Golovanvsky, M. Kang, “ Measurement and analysis of video streaming performance in live UMTS

networks," In Proceedings of WPMC'06 Conference, San Diego, CA, USA, pp: 1-5, Sept. 17-20, 2006.

[7] J.-Y. Chang, and H-L Chen," Dynamic-grouping bandwidth reservation scheme for multimedia wireless networks," IEEE Journal on Selected area in Communications, vol. 21, no. 1, pp. 1566 – 1574**,** 2003.

[8] G-R. Kwon, S-H. Park, J-W. Kim, and, S-J. Ko," Real-time R-D optimized frame-skipping transcoder for low bit rate video transmission," 6th IEEE International Conference on Computer and Information Technology. Seoul, Korea, Sept. 20-22, 2006.

[9] H. Luo, M-L. Shyu, and S-C. Chen, "An end-to-end video transmission framework with efficient bandwidth utilization," In Proceedings of the IEEE International Conference on Multimedia and Expo. ICME 2004. Taipei, Taiwan, R.O.C., pp. 623-626, June 27-30, 2004.

[10] F. Yang, Q. Zhang, W. Zhu, and Y-Q Zhang, "Bit allocation for scalable video streaming over mobile wireless Internet," Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies**.** pp: 2142-2151, March 7-11, 2004.

[11] H. M. Aziz, H. Grahn, and L. Lundberg, "Eliminating the freezing frames for the mobile user over unreliable wireless networks," In the Proceedings of the 6th International Conference on Mobile Technology, Applications and Systems. ACM Mobility Conference, France, Nice, Sept. 2- 4, 2009.

[12] M. Ries, O. Nemethova, and M. Rupp, "Performance evaluation of mobile video quality estimators," In Proceedings of the European Signal Processing Conference, EUSIPCO`07. Poznan, Poland, pp: 159-163, Sept. 3-7, 2007.

[13] T. Nguyen, P. Mehra, and A. Zakhor, "Path diversity and bandwidth allocation for multimedia streaming," In Proceedings of the International Conference on Multimedia and Expo. ICME `03. Baltimore, Maryland, pp: 1-4, 2003.

[14] International Telecommunication Union. Methodology for the subjective assessment of the quality of television pictures. ITU-R, Rec. BT.500-11, 2002.

CHAPTER SIX

Eliminating the Effect of Freezing Frames on User Perceptive by Using a Time Interleaving Technique

Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, and Lars Lundberg

Abstract

Streaming video over a wireless network faces several challenges such as high packet error rates, bandwidth variations, and delay, which could have negative effects on the video streaming and the viewer will perceive a frozen picture for certain duration due to the loss of frames. In this work, we propose a Time Interleaving Robust Streaming (TIRS) technique to significantly reduce the frozen video problem and provide a satisfactory quality for the mobile viewer. This is done by reordering the streaming video frames as groups of even and odd frames. The idea of streaming the video in this way is to avoid the losses of a sequence of neighbouring frames in case of long sequence interruption. We evaluate our approach by using a user panel and Mean Opinion Score (MOS) measurements; where the users can observe three levels of interruption to the video stream. The results show that our technique significantly improves the smoothness of the video on the mobile device in the presence of frames losses over unreliable network while the transmitted data is only increased up to 9 % (due to reduced time locality).

Keywords

Streaming video; frozen pictures,interleaving; switching frames; MOS.

6.1 Introduction

With the rapid development of video coding and wireless communication technologies, video streaming becomes very popular to mobile users (Jianhua and Jianfei, 2005). The H.264/AVC coding standard as has been introduced to achieve high compression. H.264/AVC is a lossy video compression system that removes subjective redundancy, i.e. elements of the video sequence that can be removed without significantly affecting the viewer's quality (Richardson, 2003).

The User Datagram Protocol (UDP) is a transport protocol often used for streaming video. UDP does not guarantee packet delivery; the receiver needs to rely on another protocol like Real Time Protocol (RTP) to detect packet loss (Dapeng et al., 2001). RTP is a standard protocol used over UDP for streaming videos and it is designed to provide end-to-end transport functions for supporting real-time applications. When the network is congested, the UDP sender continues to send packets at a constant rate resulting in a number of unavoidable losses. As a result, these packet losses between the video sender and the receiver significantly degrade the quality of the received video (Argyriou and Madisetti, 2005).

Real time video streaming is particularly sensitive to frame losses and delay while the video frames must arrive at the mobile device before its playout time with enough time to decode and display the contents of the video frame (Cranley and Davis, 2007). The available bandwidth in a wireless channel varies with time, and that can potentially result in lost packets or packets arriving much later than needed or temporary loss of communication, called outage (Apostolopoulos, 2001; Schierl et al., 2005).

Outage could occur to the mobile device for a creation period of time results of that the receiver buffer will be leak out of data at a constant rate. If the buffer runs out of packets, the playout of the video will be interrupted until the buffer starts receiving new packets that can guarantee enough video frames to display it on the mobile screen (Lei et al., 2007). Too long outage (e.g. several seconds) will result in user dissatisfaction; therefore outages should be maintained at an acceptable level to satisfy the Quality of Service (QoS) (Lei et al., 2007; Mancuso et al., 2004). Even a low amount of frame losses can result in a severe degradation of quality as perceived by the users (Szymanski and Gilbert, 2009; Lo et al., 2005). Streaming video over wireless networks with high error rates will affect the video quality and the viewer perceives frozen pictures for a certain duration followed by a more or less abrupt change in the picture content due to frame losses (Ong et al., 2009; Quan and Ghanbari, 2008).

In order to handle this problem the Time Interleaving Robust Streaming (TIRS) approach is proposed in this work. The purpose of TIRS is to eliminate the frozen pictures by avoiding a sequence of neighbouring frames to be lost and allowing at least every second frames to be present in the mobile device. TIRS is applied before the video stream is compressed by the H.264 encoder. The H.264 will compress the video by removing (subjectively) redundant data. Compress the video stream H.264 benefits from time locality, i.e. the difference between two adjacent frames is often relatively small.

In case of interleaving the relative difference between two adjacent frames in the video stream will increase, thus reducing time locality and increasing the amount of the data needed to represent the video. Therefore the size of the streaming video may increase due to interleaving. Consequently, interleaving will enhances the user perceived quality by reducing the problem with frozen pictures, but on the other hand it will increase the amount of transmitted data since the time locality is reduced.

Due to interleaving, adjacent frames in the original video will not arrive at the receiver after one and other, i.e. there is a time delay between the arrivals of adjacent frames. Therefore, the frames will be stored in a buffer at the receiver side before they can be played out to the display in the original order. In this case we need to introduce extra delay time for interleaving purpose.

In this paper we quantify the gain in user perceived quality caused by interleaving frames for video transmission over wireless channels with potential frame losses. We quantify the increased sized of the data representing the video due to reduced time locality. The buffer technique will on average handle network outages that are longer than the jitter buffer at the receiver side.

6.2 Background and Related Work

Frame interleaving is used in video streaming to reduce the effects of a sequence of frame losses. The sender reorders the frames before transmitting them to the receiver, so that originally adjacent frames are separated by a distance that may vary over time. Interleaving will reduce the effect of frame losses by dispersing the occurrence of errors in the original stream (Chen et al., 2004; Nafaa et al., 2008; Claypool and Zhu, 2003).

Cai and Chen (Cai and Chen, 2001) proposed an interleaving approach to improve the performance of video streaming over unreliable networks. The interleaving is applied to the compressed bitstream before the channel coding. The channel coding used was Reed-Solomon to generate channel coded block symbols and transmit these blocks over the network with the consideration of packet losses which will cause interruption at the receiving side. The packet loss may cause damages on a channel coding block so that the error recovery capability of the channel coding may be exceeded. The authors proposed the forward error correction (FEC) based pre-interleaving error control scheme for video streaming over unreliable networks. They declare that the proposed pre-interleaving greatly improved the performance of video streaming with the packet losses on ATM networks.

Schierl et al. (Schierl et al., 2005) presented a streaming system that utilizes the interleaved transmission for real-time H.264/AVC video and audio in wireless environments. In their approach they consider the audio to be the highest priority. The interleaved transmission is carried out by using priority based scheduling (PBS) with client feedback about the current fill level of each priority class in the client buffer and for retransmission on the link layer for different error rates.

The error control protocol for robust MPEG-4 video multicast over wireless channels, called unequal interleaved forward error correction (FEC) is proposed by Nafaa et al. (Nafaa et al., 2004) that can be done by combining the best features of MPEG-4 data partitioning, FEC, and interleaving techniques. The adaptive FEC transmission based on efficient feedback for 1-to-n multicast communication scenario is also considered in their study.

Tsai et al. (Tsai et al., 2009) proposed a technique to disperse the burst losses to different FEC blocks. When sending the data packets of FEC blocks over multiple paths, the proposed technique changes the transmission order of FEC blocks and sends them using path interleaving. The receiver has a packet buffer to absorb the impact of packet disordering. The path interleaving aims at striping two or more FEC blocks to multiple paths in order to share the burst packet losses into different blocks.

The above authors used a feedback technique and retransmission mechanisms for error recovery for the lost packets, the retransmission to the lost packets will increase the server overhead of fetching the needed frames from the storage and sending them to the mobile client over the wireless network. The overhead will be on the server side as well as on the wireless network and will depend on the network condition and the amount of packets has lost. Therefore we proposed the TIRS technique to avoid the use of feedback technique and retransmission mechanism.

6.3 Interleaving Distance Algorithm

Claypool and Zhu (Claypool and Zhu, 2003) proposed a video interleaving as a repair technique to ameliorate the effect of frame losses from the streaming video. They re-sequence the video frames in the video stream at the sender side, and stream them through the wireless network to the receiver side. They proposed an interleaving distance algorithm (IDA) to the recorded television program before it is encoded by using MPEG-1 and the decoding tools was used is Berkeley MPEG-2 player. In their work, the IDA for 2 and 5 is chosen, as 2 for short distance interleaving and 5 for long distance interleaving to distribute a big gap in the stream caused by the consecutive loss or propagated loss will be spread out into several small gaps.

The Group of Pictures (GOP) length (Claypool and Zhu, 2003) is used as a basis for their interleaving technique. This mean that for a GOP length for 2 distance interleaving of 9 frames per GOP (used in their paper), they would first encode the first 9 even frames (2,4,6,8,10,12,14,16,18) in the original video, and then they will encode and transmit the first 9 odd frames (1,3,5,7,9,11,13,15,17) in the video. After that they will continue with frames 20,22,24,26,28,30,32,34,36 and then frames 19,21,23,25,27,29,31,33,35, and so on. The same mechanism is applied for 5 distance interleaving of 9 frames per GOP.

The available bandwidth of wireless networks should be managed in an efficient manner. In Section 6 we investigate the code size of the IDA for the streaming video that been compressed by H.264 encoder by applying the IDA distance for 2, 3, 4, and 5 to the video frames sequence, as shown in Figure 1.

The number of frames are consider is 30 frames per seconds, where the amount of frames in each interleaving distance is equal to each other's within the same group except the interleaving distance 4 as shown in Figure 1 (d), where the amount of first odd are 8 frames followed by 8 even frames followed by 7 odd followed by 7 even frames and so on.

The interleaving mechanism is used to change the frame positions in the frames sequence based on the IDA before it is compressed and streamed over the wireless network. This can be done by splitting the video frames in the streaming server and compress them by using an H.264 encoder. The numbers of buffers needed to hold the split frames is equal to the number of interleaving distance. If the interleaving distance is $\boldsymbol{N}$, where $\boldsymbol{N} = \{2,3,4,5\}$, then the number of buffers needed is $\boldsymbol{N}$ for the server and for the client as well, as shown in Figure 2.

Frames in the video sequence will be split according to the interleaving distance $\boldsymbol{N}$. If $\boldsymbol{N}= 3$, as shown in Figure 1 (c), the frames will be forwarded to 3 different buffers. The first buffer will hold the frames number 1,4,7,10,13,16,19… , the second buffer will hold the frames number 2,5,8,11,14,17,20…, and the third buffer will hold the frames number 3,6,9,12,15,18,21… . The switching between buffers will take place to create the new frames order in the streaming video and send it to the H.264 ffmpeg encoder (ffmpeg.org) for compression.

In the normal situation, the mobile device will start receiving the video frames sequence then it will be decoded and played on the screen. For the interleaving situation the frames will arrive in different positions according to the IDA order. When the mobile devices start receiving the interleaving stream video, the video will decode the incoming frames by H.264 ffmpeg decoder (ffmpeg.org) and the frames will be split according to the IDA that are applied in the streaming server. The receiving frames will forward to the buffers that are related and a switching mechanism between buffers will applied to return the streaming sequence to its original position order as shown in Figure 2.

The latency time for playing the video will be different and based on the distance number are used, as the waiting time will be $((N-1)/N) * Ts$, where $\boldsymbol{N}$ is the interleaving distance, and $\boldsymbol{Ts}$ is the time in seconds. As an example, if we assume that the IDA is 3 then the interleaving time with the waiting time to start playback the video is 0.67 second given a video with 30 frames per second (see Figure 1 (c)). We can observe from that the more the interleaving distances are applied the more waiting time is needed to start playback the receiving frames by the mobile device.

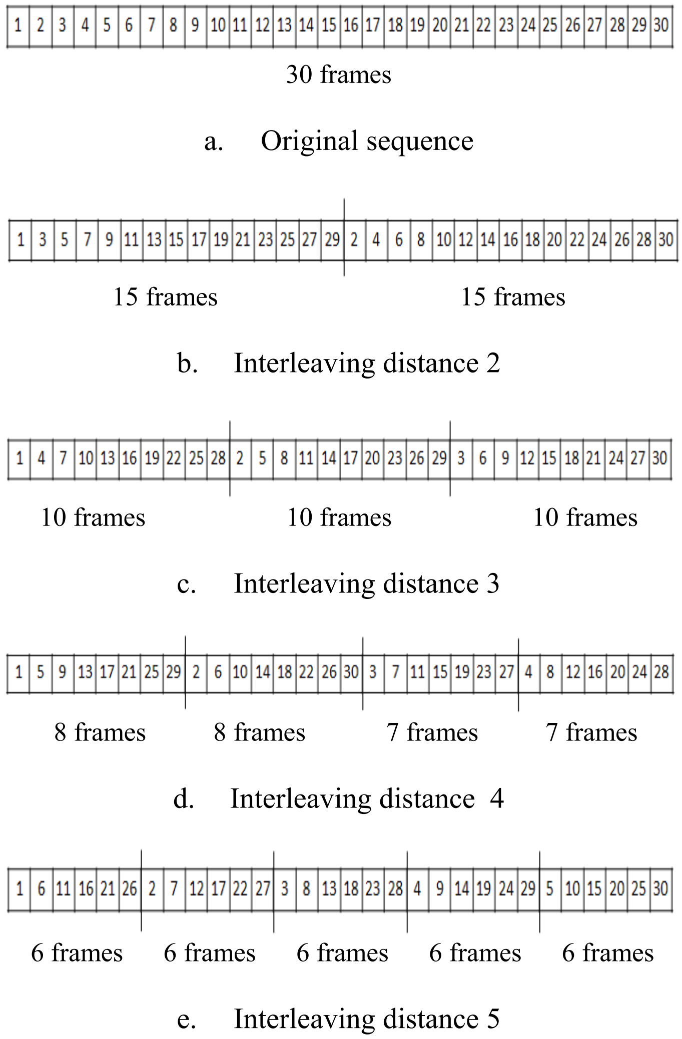


Figure 1: Frame sequences with different interleaving distances for 30 frames.

The reconstruction approach (Liu and Claypool, 2000) used to recover the lost frames is based on repeating the previous consecutive frame. Their reconstruction approach is done before the transmission, the encoder will generate two version of compressed frames, one with high quality (primary frame) and the another one with low quality (secondary frame). If the primary frames are received successfully then the decoder will discard the secondary frame. If the primary frame is lost the decoder will wait for the next package that carries the secondary frame to extract. When both the primary and the secondary frame are lost then the frame cannot be repaired.

The main limitation in their work is that if the primary frame is lost a delay could occurs for waiting for the secondary frame to be extracted and played at the main time the jitter buffer will be underflow. While the overhead will be increased and depend on the percentage of the redundancy are been implemented. Another important limitation on their work is that if there is a long sequences of frames are lost in case of outage then a sequence of frames of primary and secondary frames will be not be available in the decoder and the user will experience a frozen pictures. Therefore a Time Interleaving Robust Streaming (TIRS) technique is proposed in this work to overcome the limitation in above and to stream the video frames without transmitting any redundancy frames.

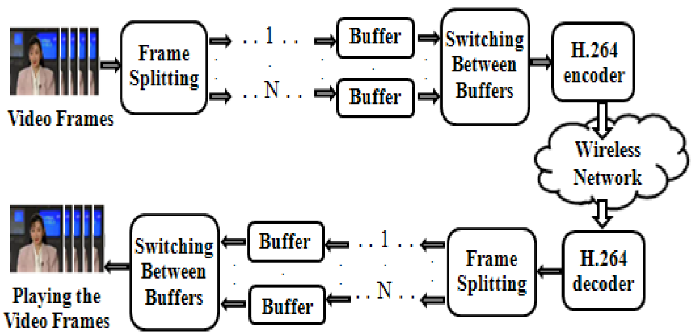


Figure 2: Streaming video as IDA over wireless network.

6.4 Time Interleaving Robust Streaming

In this section, we propose a Time Interleaving Robust Streaming (TIRS) technique to stream the video frames. The idea behind the TIRS technique is to avoid frozen pictures in the mobile screen over unreliable network. This can be done by streaming the video as a group of even and odd frames. The TIRS will be applied before the video has been encoded and compressed by the H.264 encoder and in this study the time interleaving (Δt) is considered for 1, 2 and 3 seconds. The TIRS can be implemented for different interleaving times to distribute the big lost of frames on different positions in the streaming sequence and to avoid a sequence of interruption to the streaming video.

In following example the Δt = 1s. The mechanism is applied by splitting the frames to even frames follow by odd frames and if we assume that the number of frames 30 frames/second and we refer to it as a frame group (FG), where Fe and Fo are groups of even and odd frames belonging to the same FG as shown in Figure 3.

The second stage is to create the streaming video based on the interleaving technique, where the even and odd frames will be grouped and distributed in different position in the streaming sequence. The distribution mechanism is depending on the system design either Fe followed by Fo or Fo followed by Fe.

The even frames (Fe) from FG-1 and FG-2 will be grouped and streamed followed by a sequence of odd frames (Fo) from FG-1, FG-2, FG-3 and FG-4, and followed by a sequence of even frames (Fe) from FG-3, FG-4, FG-5 and FG-6 and so on, and as shown in Figure 4.

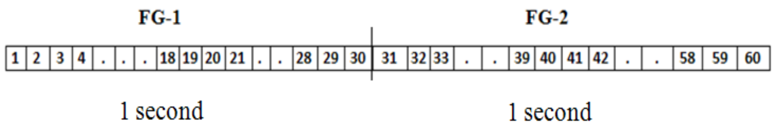
The reason why we send 4 groups of odd frames instead of 2 groups is that we want to minimize the reduction of time redundancy. This will increase the compression rate when using H.264.

The proposed TIRS technique is taking place in the streaming server, where the sequence of groups of even and odd frames is split and queued in two different buffers, where the even buffer contains the Fe's frames and the odd buffer contains the Fo's frames.

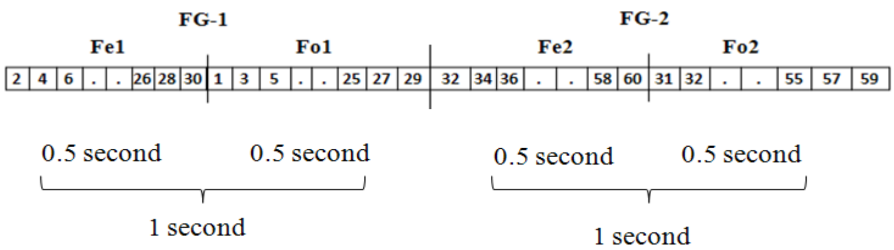
To implement the TIRS technique, a switching between buffers is used to switch between groups of even and odd frames according to the time interleaving parameters been set and before its been encoded by H.264 encoder after that, the time interleaving streaming sequence will then be streamed through the wireless network as shown in Figure 5. When the mobile device starts receiving the streaming video, the H.264 decoder will decode the video and the frames will be split into even and odd frames and forwarded to two buffers. The even buffer will hold the even frames, while the odd buffer will held the odd frames.

The even frames will be delayed by Δt waiting for the odd frames to arrive to the odd buffer. The number of frames that has arrived to the even buffer is $V = \Delta t.R$, where R is the number of frames per second received by the buffer.

Fig 6 (a) shows how the frames are stored in the even buffer, and Fig 6 (b) shows the odd buffers. At time Δt the playout of the video starts. Fig 6 (a) and (b) show that after the initial startup, i.e after time Δt, the sum of the number of frames in the two buffers is always V.



a. Original stream



b. Frames interleaving within the same FG, which is similar to the IDA technique

Figure 3: Video streaming sequence.

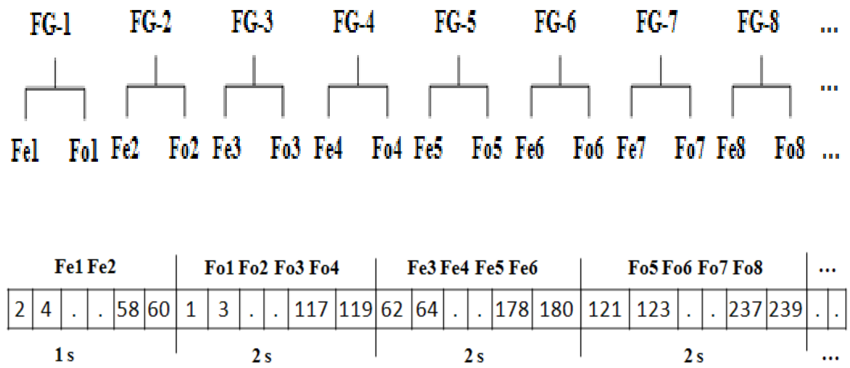


Figure 4: The proposed TIRS technique, where Δt = 1s.

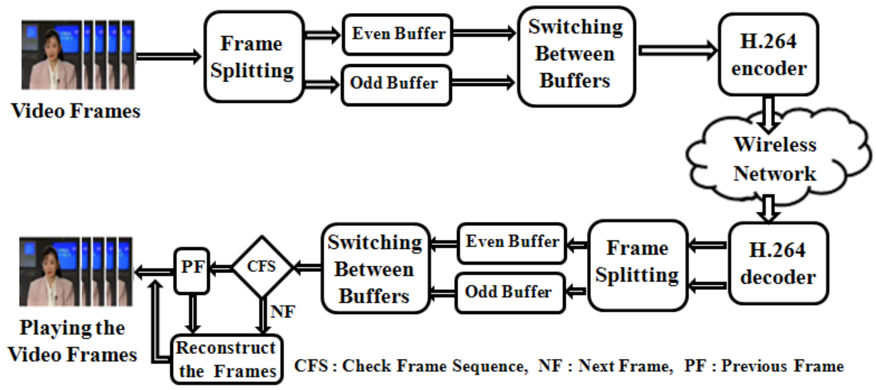


Figure 5: Streaming video using TIRS over wireless networks.

The check frame sequence (CFS) will be used to check the missing frames from the streaming video (Aziz et al., 2010). If there is a missing frame a reconstruction mechanism will be applied to create the missing frame based on linear interpolation between the next and previous frames (Peng et al., 2005).

6.5 The Effect of Losses on the Interleaving Frames

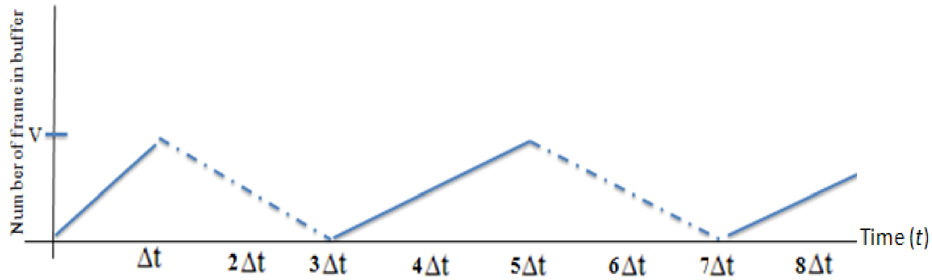
Streaming video over unreliable network could have a significant effect on the video quality, especially when there is a consecutive sequence of frames lost. Mobile device employ playback buffers where video frames are stored before being displayed.

The playback of a video starts when there are available frames in the buffer, in the TIRS case compensate for the delay due to frame group (FG) interleaving. Interleaving has the advantage that we can use frame interpolations to tolerate loss of frame sequences up to a certain length. By using long interleaving intervals (e.g. Δt =2s instead of Δt =1s) we can obviously tolerate loss of longer frame group sequences, which going to be helpful in the case of outages. On the other hand, long interleaving intervals means that the delay due to buffering gets longer, which is a disadvantage in many cases, e.g. when looking at live events and when changing from one channel to another when watching TV. Consequently, we would like to have short buffer delays while at the same time being able to tolerate losses of long frames sequence (long outages).

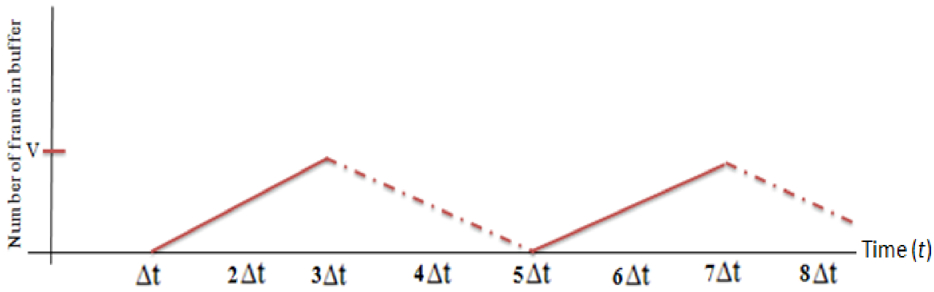
Figure 6 (c) shows $f(t)$ which is the maximum outage length that we can handle at time t given that no frames outside of the outage are lost; t = 0 marks the start of the video. The delay buffer due to frame interleaving in as described in Fig 4, Δt =1s. This means that the video will be shown at the receiver at least at Δt after it has been transmitted from the sender. The sender will start by sending even frames based on the value of Δt and then it will send odd frames for $2\Delta t$ and then even frames for $2\Delta t$, and then odd frames for Δt, and so on.

- If the outage starts at t = 0 until Δt, then all the even frames are lost and if the odd frames are received by the mobile device (assuming that no additional frames are lost), then the lost even frames can be recreated by interpolating the odd frames from Δt to $2\Delta t$.

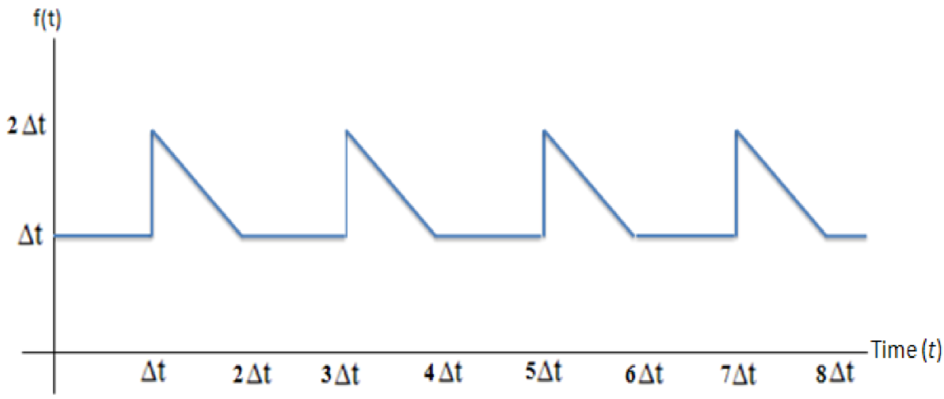
- If the outage starts at Δt until $2\Delta t$, then the lost odd frames can be recreated by interpolating the receiving even frames from 0 to Δt.



a. Buffer receiving the even frames



b. Buffer receiving the odd frames



c. The effect of frame losses on the interleaving stream

Figure 6: The buffering behaviour corresponding to the frame types and frame lost.

- If the outage starts at t = 0 until 1.5Δt, then the outage duration is longer than Δt and in this case we are facing a frozen video.

We can observed from that, if the outage occur in the initial of the streaming video and its greater than Δt, then we will face a frozen video but if the outage occurs between Δt and $2\Delta t$ then there is possibility to interpolate the missing frames. If the outage is occurs between two interval as an example between the second half between Δt to $2\Delta t$ and the first half between $2\Delta t$ and $4\Delta t$ and its greater than Δt in this case we will face a frozen picture.

6.6 The Effect of Interleaving on the File Streaming Size

Compressing the video frames based on IDA and TIRS technique by using H.264 ffmpeg encoder (ffmpeg.org) will affect the amount of redundancy to be removed. As the video frame positions are changed and similar redundancies are transmitted which lead to increase the size of the video stream for both techniques.

The IDA and TIRS techniques have been applied on the video test sequences Akiyo, Foreman, News, and Waterfall (trace.eas.asu.edu), with a transmission rate of 30 frames per second, and each of 10 seconds duration with QCIF (176×144 pixels) resolutions. The chosen videos are well known as professional test sequence. The chosen videos have different characteristics and different motions. Therefore the sizes of the interleaving stream are expected to be different from one video to another.

6.6.1 The Effect of IDA on the File Streaming Size

Implementing the interleaving distance algorithm (IDA) to the video frames and compressed by H.264 ffmpeg encoder will affect the redundancy to be removed. The frames position are changed according to the chosen distance algorithm 2,3,4, or 5, where similar redundancy is transmitted several times which lead to increase the size of the video.

As shown in Table 1, the sizes of the compressed videos clips are increased differently for different interleaving distances. The reason for this is that the frames are moved to different scene positions which affects the time redundancy and thus also the ability to compress the video using H.264.

For the Foreman video clip, the size is increased less than the rest of the testing videos as the Forman clip is based on two main scenes. The 1st scene where Foreman moving his head and pointing to the construction side, while the 2nd scene is more constant all the time. Therefore, the frames will be within the two main scenes, where the 2nd scene is less motion than the first one and therefore the size of the video is not that much effected.

Table 1: The compressed video size (bytes) and the number of packets for IDA.

Videos	**Akiyo**	**Foreman**	**News**	**Waterfall**
Original	2886120	7151874	4115786	6679052
No. of packets	1961	4859	2796	4537
IDA-2	3112270	7356044	4478530	7283722
No. of packets	2114	4997	3042	4947
Increase	**7.8%**	**2.8%**	**8.8%**	**9.0%**
IDA-3	3240352	7419332	4648352	7482042
No. of packets	2201	5040	3158	5083
Increase	**12.2%**	**3.7%**	**12.9%**	**12.0%**
IDA-4	3314786	7432886	4735802	7622492
No. of packets	2252	5049	3217	5178
Increase	**14.8%**	**3.9%**	**15.0%**	**14.1%**
IDA-5	3329254	7406416	4716468	7627336
No. of packets	2262	5032	3204	5182
Increase	**15.3%**	**3.5%**	**14.5%**	**14.2%**

6.6.2 The Effect of TIRS on the File Streaming Size

The study of the effect of different time interleaving on the streaming file size, where the value of Δt are considered is 1, 2 and 3 seconds. For Δt = 1 second interleaving the initial stream is based on streaming the even frames from 2 groups, followed by odd frames from 4 groups and so on as been shown in Figure 4. For Δt = 2 second interleaving the initial stream is based on streaming the even frames from 4 groups followed by odd frames from 8 groups and so on, while for Δt =3 seconds interleaving the initial stream is based on streaming the even numbers from 6 groups followed by odd frames from 12 groups and so on.

The Δt = 1, 2 and 3 seconds interleaving been applied before it been compressed by H.264 ffmpeg encoder (ffmpeg.org), and the sizes of the compressed videos clips are increased differently and for different interleaving time than the original clip as shown in Table 2.

The size of the test videos are slightly increased due to the change of the frames positions, as the group of sequence frames are moved from one different scene to another. The data of the moved frame is different from the new scene, therefore the H.264 ffmpeg encoder will compress extra redundant data for a group of frames that are moved from their original scene position to different one.

Where the changes in the streaming file size for the three interleaving time are mostly similar to each other except the News clips as the 2 second interleaving show less increases in the streaming size this due to the movement of the group of frames to a position where there is similar data and the encoder will remove more redundant data than the others.

Table 2: The compressed video size (bytes) and the number of packets for TIRS.

Videos	**Akiyo**	**Foreman**	**News**	**Waterfall**
Original	2886120	7151874	4115786	6679052
No. of packets	1961	4859	2796	4537
Δt=1	**3124282**	**7365602**	**4498224**	**7237768**
No. of packets	2122	5004	3056	4916
Increase	**8.2%**	**2.9%**	**9.2%**	**8.3%**
Δt=2	**3122252**	**7379988**	**4261158**	**7238698**
No. of packets	2121	5013	2895	4918
Increase	**8.1%**	**3.1%**	**3.5%**	**8.3%**
Δt=3	**3121174**	**7364790**	**4491916**	**7234486**
No. of packets	2120	5003	3052	4915
Increase	**8.1%**	**2.9%**	**9.1%**	**8.3%**

6.7 Comparison between IDA and TIRS Technique

The compressed size of the video stream will be increased when implementing the IDA and the TIRS technique. The increases of the video size will be different from one video to another and based on the moving position of the frames.

For the IDA, the compressed size of the video will increase as the interleaving distance is increased, the frames could possible move between surrounding frames that have different data. The compressed sizes of our proposed TIRS technique will increase slightly and depend on the group movement.

For short interleaving e.g. IDA-2 and Δt =1s in TIRS the size increase is more or less the same for IDA and TIRS. For long interleaving, that can handle long outages, e.g. IDA-5 and Δt =3s in TIRS, the size increase is longer in IDA.

The numbers of buffers are needed for the IDA will depend on the number of distance interleaving (***N***) that are considered in the system design. While for the TIRS approach we only used 2 buffers, therefore our proposed approach will reduce the complexity in the interleaving process.

The initial waiting time for playback the video for IDA is increased as the interleaving distance (***N***) is increases, while for our proposed TIRS approach the waiting time will dependent on the duration of the interleaving time, as more the duration time as more the waiting time is going to be to playback the video on the mobile device.

The outage length that can be tolerate by using IDA its less than TIRS, as TIRS can handle the outage length based on the time interleaving duration as higher the value of Δt as longer the outage length can be tolerate.

6.8 Subjective Viewing Test

6.8.1 Testing Methods

It is well known that peak signal-to-noise ratio (PSNR) does not always rank quality of an image or video sequence in the same way as a human being. There are many other factors considered by the human visual system and the brain (Martinez-Rach et al. 2007).

One of the most reliable ways to assess the quality of a video is subjective evaluation using the Mean Opinion Score (MOS). MOS is a subjective quality metric obtained from a panel of human observers. It has been regarded for many years as the most reliable form of user perceived quality measurements (Martinez-Rach et al. 2008).

The MOS measurements are used to evaluate the video quality in this study and follow the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). The score grades in this method are ranged from 0 to 100 which are mapped to the quality ratings on the 5-grade discrete category scale labelled with Excellent (5), Good (4), Fair (3), Poor (2) and Bad (1). The physical lab, with controlled lighting and set-up, conforms to the ITU-R recommendation (ITU-R, 2002).

The subjective experiment has been conducted at Blekinge Institute of Technology in Sweden. The users observed the video clip with frozen picture and the proposed TIRS technique, with the participation of 30 non-expert test subjects, 26 males and 4 females. They were all university staff and students and within an age range from 23 to 37.

6.8.2 Testing Materials and Environments

The study is done by coding the video test sequences Akiyo, Foreman, News, and Waterfall (trace.eas.asu.edu), and compressed the video by using the H.264 ffmpeg codec (ffmpeg.org) for the proposed interleaving technique. The chosen videos are coded as 25 frames/second with a resolution of 176 x 144. The transmission rate is 30 frames/second and the number of frames transmitted is 1800.

The outage duration time are set for this study is 650, 1300 and 2000 milliseconds for short, medium, and long respectively. The video sequences are shown on a 17 inch EIZO FlexScan S2201W LCD computer display monitor with a native resolution of 1680 x 1050 pixels. The video sequences for the frozen and the proposed technique are displayed with a resolution of 176 x 144 pixels in the centre of the screen with a black background with a duration of 60 seconds for each video sequence.

6.9 Experimental Results

The video quality is subject to the personal opinion; where the quality of service improvements for video transmission has the only goal to satisfy the average human watching the contents of the video stream. For the frozen technique, the observer evaluates the video stream, where the effect of the frozen pictures on the video streaming based on the missing frames due to the lost packets during the outage case. While for the proposed TIRS technique, the observers evaluate the video based on Δt =1 second, where the missing frames been constructed based on the surrounding frames. Therefore the interpolation between the available surrounding frames will be applied to create the missing frames in the mobile devices.

The data are gathered from the two subjective experiment groups and are based on the mean opinion scores rating from Excellent to Bad that have been given by the individual viewers for evaluating each video clip. A concise representation of the data can be achieved by calculating conventional statistics such as the average, standard deviation, mean score and 95% confidence interval, of the related distribution of scores. The statistical analysis of the data from the subjective experiments reflects the fact that perceived quality is a subjective measure and has been described statistically as shown in Figure 8 and Table 3, where the statistical data analysis for the user's evaluation for different videos contents (VC) and for two techniques.

For the frozen technique it can be clearly seen that the observers manages to identify the frozen pictures in the testing video. For the short handover time, the MOS is lower than 3 for the testing videos except the Akiyo testing video is lower than 3.5 corresponding to the five-level quality scale ranks, where Akiyo video is less motion than the rest of the testing video. For medium and long handover time, the MOS is lower than 3, due to the higher percentage of the missing frames, which the viewer easily notice the frozen pictures.

The MOS for our proposed TIRS technique, it is show that the observer rank the video test larger than 3 for the short handover time, except the Akiyo video is larger than 3.5.

For medium and long handover time, the MOS is larger than 3 for the testing video except the News video is larger than 2.5 as shown in Figure 8 (b), (c), due to the effect of the reconstruction mechanism for higher frames been missing in the decoder, which can easily been notice by the observers.

The rating scores for the user evaluation and according to the rating category scales that been given, where **r** denote the vector of the number of ratings in each category provided by the users in a certain experiment, i.e., $\boldsymbol{r} = [r_{\text{Bad(1)}}, r_{\text{Poor(2)}}, r_{\text{Fair(3)}}, r_{\text{Good(4)}}, r_{\text{Excellent(5)}}]$. For each anticipated loss ratio $\boldsymbol{l}$, we obtain one vector of ratings $\mathbf{r}^{l}$ without taking advantage of interleaving, and a corresponding vector $\mathbf{r}^{l,\text{TIRS}}$ when using the TIRS method. The difference vector $\Delta r = \boldsymbol{r}^{l,\text{TIRS}} - \boldsymbol{r}^{l}$ illustrates the change of the numbers of user rankings per category: if $\Delta r_i < 0$, then the number of corresponding user ratings in category $\boldsymbol{i}$ has decreased when employing the TIRS technique, and vice versa. If TIRS yielded an improvement, one would actually expect a decrease in the number of negative rankings ($\Delta r_{\text{Bad(1)}} \leq 0$, $\Delta r_{\text{Poor(2)}} \leq 0$) and on the other hand a growth in the number of positive rankings ($\Delta r_{\text{Good(4)}} \geq 0$, $\Delta r_{\text{Excellent(5)}} \geq 0$). A negative staple means that the corresponding ranking has diminished, while a positive staple means that there are high rankings as compared to the frozen technique as showing in Figure 9. As been shown clearly that the number of observers rating our proposed TIRS techniques higher than frozen techniques and for each category, the rating goes under zero, indicate that the observer give high rating for the Bad (1) and Poor (2) category for frozen technique, while the rating goes upper than zero, indicate that the observer give high rating for Good (4) and Excellent (5) for our TIRS technique and low for the frozen technique. Figure 8 shows that the biggest change due to the TIRS technique are in the categories Poor (2) and Good (4), whereas the change in the categories Bad (1) and Excellent (5) is more limited.

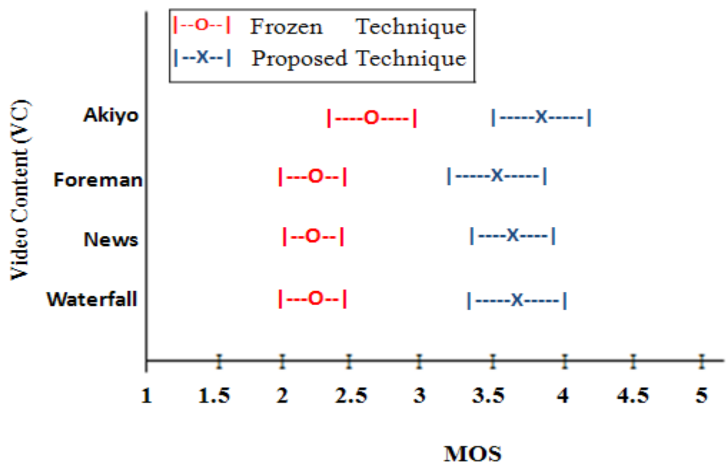
It can be observed from that the videos presented to the viewers resulted in a wide range of perceptual quality ratings indeed for both experiments. When we analyse their scores we feel that our proposed TIRS technique is a satisfactory mechanism to distribute the missing frames in the streaming video to eliminate the frozen pictures with the reconstruction to the missing frames and provide a smooth video in the mobile screen.

6.10 Conclusions

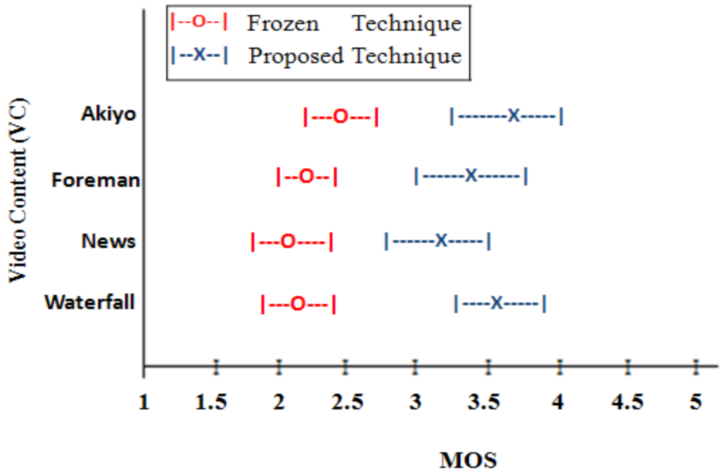
Wireless network transmission errors in forms of frame losses could have a major impact on the end user experience of real time video streaming. A Time Interleaving Robust Streaming (TIRS) technique is proposed to minimize the effect of the outage time on the frames sequence. Streaming the video frames according to the TIRS technique makes it possible to achieve this and to avoid a sequence of consecutive video frames to be lost during the transmission.

The advantage of TIRS is that the lost of the frame sequences will spread out on the streaming video with the ability to be reconstructed at the receiver side. This means that if 30 consecutive frames are lost we do not get a 1 second freezing (assuming 30 frames/second); as 30 frames are distributed to 2 frame groups related to another 2 receiving groups. Because of this, the lost frames where can be interpolated by using the surrounding frames. The disadvantage of interleaving is that the size of the video stream is increased slightly (less than 10%).

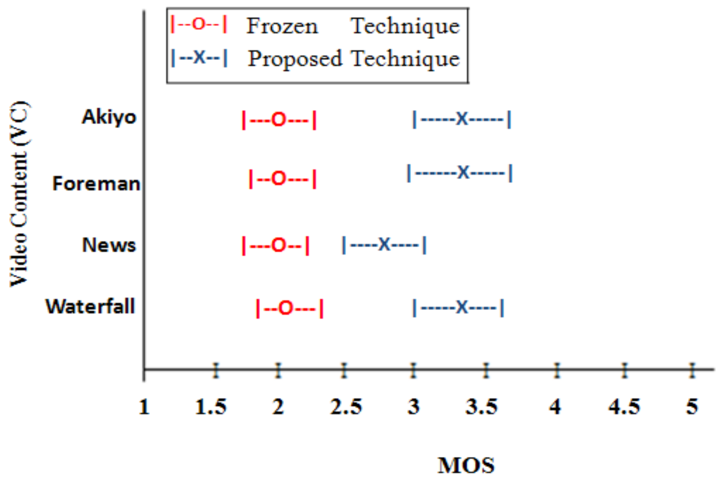
Our analysis based on user panel tests and showed that there is a significant performance improvement for video smoothness for different outage time.



a. MOS for short outage time.



b. MOS for medium outage time.



c. MOS for long outage time.

Figure 8: The MOS for different videos content and for different outage time.

Table 3: The statistical data analysis for different videos and for different outage time.

a. Statistical data analysis for short outage time.

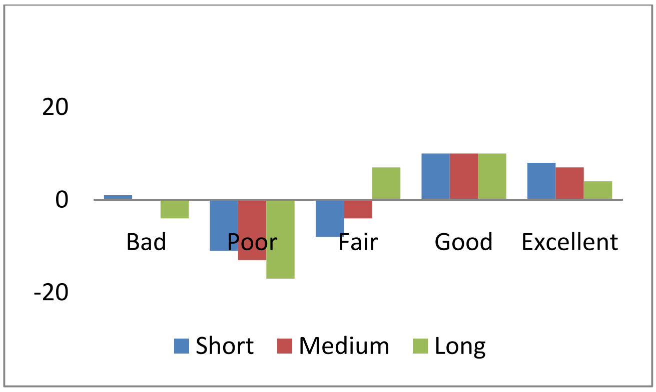
	Frozen Technique				**Proposed Video**			
	Akiyo	**Foreman**	**News**	**Waterfall**	**Akiyo**	**Foreman**	**News**	**Waterfall**
Avg	2,70	2,27	2,27	2,27	3,87	3,57	3,67	3,70
StD	0,65	0,52	0,45	0,52	0,97	0,90	0,66	0,92
CoV	24%	23%	20%	23%	25%	25%	18%	25%
95 % CI	9,0%	8,5%	7,4%	8,5%	9,3%	9,4%	6,7%	9,3%

b. Statistical data analysis for medium outage time.

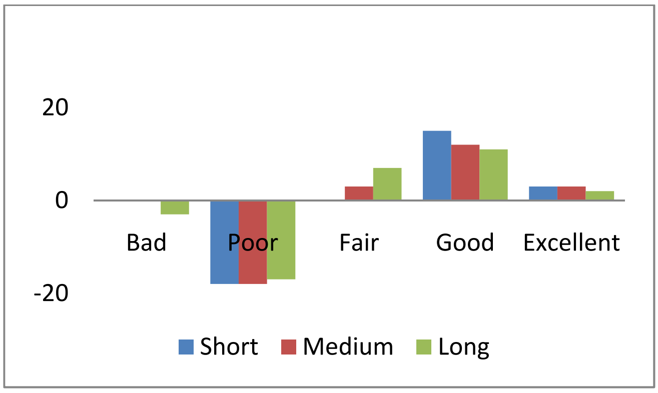
	Frozen Technique				**Proposed Video**			
	Akiyo	**Foreman**	**News**	**Waterfall**	**Akiyo**	**Foreman**	**News**	**Waterfall**
Avg	2,40	2,20	2,03	2,17	3,63	3,40	3,13	3,63
StD	0,56	0,48	0,61	0,53	1,07	0,97	0,97	0,96
CoV	23%	22%	30%	24%	29%	29%	31%	26%
95 % CI	8,7%	8,1%	11,2%	9,1%	11,0%	10,6%	11,5%	9,8%

c. Statistical data analysis long outage time.

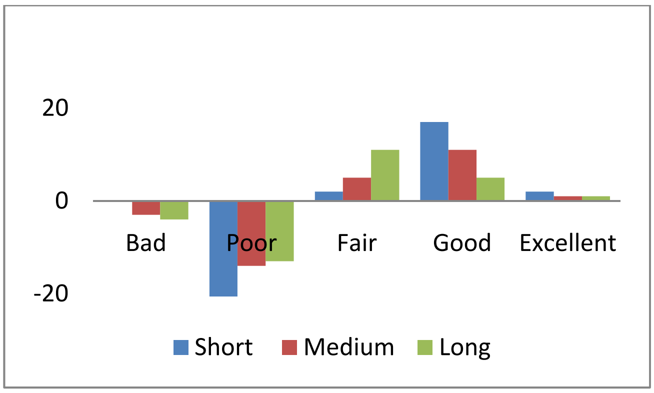
	Frozen Technique				**Proposed Video**			
	Akiyo	**Foreman**	**News**	**Waterfall**	**Akiyo**	**Foreman**	**News**	**Waterfall**
Avg	2,00	2,03	1,93	2,10	3,43	3,30	2,87	3,37
StD	0,59	0,56	0,58	0,55	0,97	0,92	0,90	0,96
CoV	30%	28%	30%	26%	28%	28%	31%	28%
95 % CI	11,0%	10,3%	11,2%	9,8%	10,5%	10,4%	11,7%	10,6%



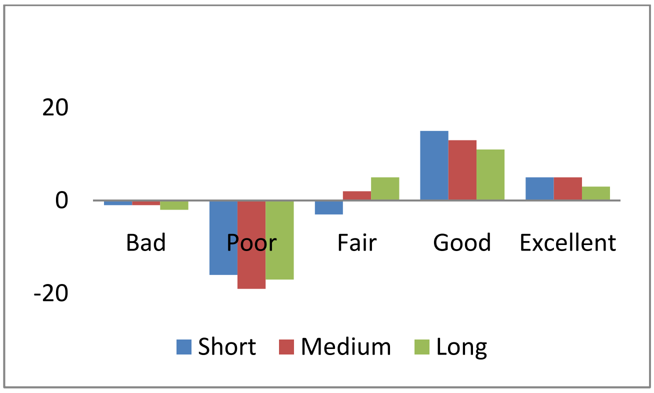
a. Akiyo video



b. Foreman video



c. News video



d. Waterfall video

Figure 9: Difference of number of user ratio per category when employing the TIRS Technique and for different outage time.

ACKNOWLEDGMENTS

We would like to thank the staff and students at Blekinge Institute of Technology, Sweden for participating in the subjective experiments. We also would like to thank the Swedish Knowledge Foundation for sponsoring a part of this work through the project QoEMoS (21601420).

References

Apostolopoulos, J., 2001. Reliable video communication over lossy packet networks using multiple state encoding and path diversity. In: Proceedings of Visual Communications and Image Processing, pp. 392-409.

Argyriou A., Madisetti V., 2004. Streaming H.264/AVC video over the internet. In: IEEE Consumer Communications and Networking Conference (CCNC’04), pp. 169-174.

Aziz H.M., Fiedler M., Grahn H., Lundberg L., 2010. Streaming video as space – divided sub-frames over wireless networks. In: the Proceedings of the 3rd Joint IFIP Wireless and Mobile Networking Conference, (WMNC'10).

Cai J., Chen C.W., 2001. FEC-based video streaming over packet loss networks with pre-interleaving. In: Proceeding for the International Conference on Information Technology: Coding and Computing, pp. 10 – 14.

Chen L.- J., Sun T., Sanadidi M.Y., Gerla M., 2004. Improving wireless link throughput via interleaved FEC. In: Proceedings of the 9th IEEE Symposium on Computers and Communications, pp. 539- 544.

Claypool M., Zhu Y., 2003.Using interleaving to ameliorate the effects of packet loss in a video stream. In:Proceedings of the International Workshop on Multimedia Network Systems and Applications (MNSA).

Cranley N., Davis M., 2007. Video frame differentiation for streamed multimedia over heavily loaded IEEE 802.11e WLAN using TXOP. In: Proceedings of 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07).

Dapeng W., Hou Y. T., Zhu W., Zhang Y.-Q., Peha J. M., 2001. Streaming video over the internet: approaches and directions. IEEE Transactions on Circuits and Systems for Video Technology, 11(3), 282-300.

International Telecommunication Union. Methodology for the subjective assessment of the quality of television pictures. ITU-R, Rec. BT.500-11, 2002.

Jianhua W., Jianfei C., 2005. Quality-smoothed encoding for real-time video streaming applications. In: Proceedings of the 9th International Symposium on Consumer Electronics (ISCE'05), pp. 445 – 449.

Lo A., Heijenk G., Niemegeers I., 2005. Performance evaluation of MPEG-4 video streaming over UMTS networks using an integrated tool environment. In: Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'05), pp. 676-682.

Liu L., Claypool M., 2000. Using Redundancy to Repair Video Damaged by Network Data Loss. Multimedia Computing and Networking (MMCN).

Lei H., Fan C., Zhang X., Yang D., 2007. QoS aware packet scheduling algorithm for OFDMA systems. In : IEEE 66th Vehicular Technology Conference, pp.1877-1881.

Martinez-Rach M., Lopez O., Pinol P., Malumbres M.P., Oliver J., Calafate, C. T., 2007. Quality assessment metrics vs. PSNR under packet loss scenarios in MANET wireless networks. In: Proceedings of the International Workshop on Mobile Video (MV '07), pp. 31-36.

Martinez-Rach M., Lopez O., Pinol P., Malumbres M. P., Oliver J., Calafate C. T., 2008. Behavior of quality assessment metrics under packet losses on wireless networks. XIX Jornadas de Paralelismo.

Mancuso V., Gambardella M., Bianchi G., 2004. Improved support for streaming services in vehicular networks. In : Proceedings of ICC 2004 conference, pp. 4362-4366.

Nafaa A., Ahmed T., Mehaoua A. 2004. Unequal and interleaved FEC protocol for robust MPEG-4 multicasting over wireless LANs. In: Proceedings of the IEEE International Conference on Communications (ICC'04), pp. 1431 – 1435.

Nafaa A., Taleb T., Murphy L., 2008. Forward error correction strategies for media streaming over wireless networks. IEEE Communications Magazine: Feature Topic on New Trends in Mobile Internet Technologies and Applications 46(1), 72-79.

Ong E. P., We S., Loke M. H., Rahardja S., Tay J., Tan C. K., Huang L., 2009. Video quality monitoring of streamed videos. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP `09), pp. 1153 – 1156.

Peng Y.-C, Chang H.-A., Chen C.-K. L., Kao. H. C.-J. Kao, 2005. Integration of image stabilizer with video codec for digital video cameras. In: IEEE International Symposium on Circuits and Systems (ISCAS'05), pp. 4871 – 4874.

Quan H.-T., Ghanbari M., 2008. Asymmetrical temporal masking near video scene change. In: Proceedings of the IEEE International Conference on Image Processing (ICIP´08), pp. 2568-2571.

Richardson I. E. G. R., 2003. H.264 and MPEG-4 video compression video coding for next-generation multimedia. John Wiley & Sons Ltd,.

Schierl T., Kampmann M., Wiegand T., 2005. H.264/AVC interleaving for 3G wireless video streaming. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME ’05), pp. 868-871.

Szymanski H. T., Gilbert D., 2009. Internet multicasting of IPTV with essentially-zero delay jitter. IEEE transactions on broadcasting 55(1), 20-30.

Tsai M.-F., Ke C.-H., Kuo C.-I., Shieh C.-K., 2009. Path dependent adaptive forward error correction with multipath interleaving control scheme for video streaming over wireless networks. In: IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 1236 – 1239.

trace.eas.asu.edu/yuv/index.html.

www.ffmpeg.org.

ABSTRACT

Real time video streaming over wireless network is an increasingly important and attractive service to the mobile users. Video streaming involves a large amount of data to be transmitted in real time, while wireless channel conditions may vary from time to time. It is hard to guarantee a reliable transmission over the wireless network, where the parameters specifying the transmissions are; bandwidth, packet loss, packet delays, and outage times. The quality of the video is affected negatively when network packets are lost, and the mobile users may notice some sudden stop during the video playing; the picture is momentarily frozen, followed by a jump from one scene to a totally different one

The main objective of this thesis is to provide a smooth video playback in the mobile device over unreliable networks with a satisfactory video quality. Three different techniques are proposed to achieve this goal. The first technique will stream duplicate gray scale frames over multichannels, if there is lost frames in one channel it can be recovered from another channel. In the second technique, each video frame will be split into sub-frames. The splitted sub-frames will be streamed over multichannels. If there is a missing sub-frame during the transmission a reconstruction mechanism will be applied in the mobile device to recreate the missing sub-frames. In the third technique, we propose a time interleaving robust streaming (TIRS) technique to stream the video frames in different order. The benefit of that is to avoid the losses of a sequence of neighbouring frames. A missing frame from the streaming video will be reconstructed based on the surrounding frames.

The mean opinion score (MOS) metric is used to evaluate the video quality. The experienced quality of a video is subject to the personal opinion, which is the only goal to satisfy the average human watching the contents of the video.

BLEKINGE TEKNISKA HÖGSKOLA · BTH ·

ISSN 1650-2140

2010:09

ISBN: 978-91-7295-190-7