STREAMING VIDEO OVER UNRELIABLE AND BANDWIDTH LIMITED NETWORKS

Hussein Muzahim Aziz

Blekinge Institute of Technology
Doctoral Dissertation Series No. 2013:08

School of Computing

Streaming Video over Unreliable and Bandwidth Limited Networks

Hussein Muzahim Aziz

Blekinge Institute of Technology doctoral dissertation series
No 2013:08

Streaming Video over Unreliable and Bandwidth Limited Networks

Hussein Muzahim Aziz

Doctoral Dissertation in
Computer Systems Engineering

School of Computing
Blekinge Institute of Technology
SWEDEN

School of Computing
Publisher: Blekinge Institute of Technology,
SE-371 79 Karlskrona, Sweden
Printed by Printfabriken, Karlskrona, Sweden 2013
ISBN: 978-91-7295-258-4
ISSN 1653-2090
urn:nbn:se:bth-00560

Abstract

The main objective of this thesis is to provide a smooth video playout on the mobile device over wireless networks. The parameters that specify the wireless channel include: bandwidth variation, frame losses, and outage time. These parameters may affect the quality of the video negatively, and the mobile users may notice sudden stops during the playout video, i.e., the picture is momentarily frozen, followed by a jump from one scene to a different one.

This thesis focuses on eliminating frozen pictures and reducing the amount of video data that need to be transmitted. In order to eliminate frozen scenes on the mobile screen, we propose three different techniques. In the first technique, the video frames are split into sub-frames; these sub-frames are streamed over different channels. In the second technique the sub-frames will be "crossed" and sent together with other sub-frames that are from different positions in the streaming video sequence. If some sub-frames are lost during the transmission a reconstruction mechanism will be applied on 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 losing a sequence of neighbouring frames. A missing frame from the streaming video will be reconstructed based on the surrounding frames on the mobile device.

In order to reduce the amount of video data that are streamed over limited bandwidth channels, we propose two different techniques. These two techniques are based on identifying and extracting a high motion region of the video frames. We call this the Region Of Interest (ROI); the other parts of the video frames are called the non-Region Of Interest (non-ROI). The ROI is transmitted with high quality, whereas the non-ROI is interpolated from a number of references frames. In the first technique the ROI is a fixed size region; we considered four different types of ROI and three different scenarios. The scenarios are

based on the position of the reference frames in the streaming frame sequence. In the second technique the ROI is identified based on the motion in the video frames, therefore the size, position, and shape of the ROI will be different from one video to another according to the video characteristic. The videos are coded using ffmpeg to study the effect of the proposed techniques on the encoding size.

Subjective and objective metrics are used to measure the quality level of the reconstructed videos that are obtained from the proposed techniques. Mean Opinion Score (MOS) measurements are used as a subjective metric based on human opinions, while for objective metric the Structural Similarity (SSIM) index is used to compare the similarity between the original frames and the reconstructed frames.

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 Professor 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 with special thanks to the students at Blekinge Institute of Technology, Sweden for participating in the subjective experiments.

I would like to thank the Swedish Knowledge Foundation for sponsoring a part of this work through the project QoEMoS (d-nr 2008/0520).

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, June 2013

Hussein Muzahim Aziz

This work is dedicated to Yosser

List of Publications

The thesis is based on the following publications:

1. Hussein Muzahim Aziz, Håkan Grahn, Lars Lundberg. Eliminating the Freezing Frames for the Mobile User over Unreliable Wireless Networks. ACM Mobility Conference. The 6th International Conference on Mobile Technology, Applications and Systems, Nice, France, September 2009, pp. 57:1-57:4. As chapter 2.
2. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Streaming Video as Space – Divided Sub-Frames over Wireless Networks. The 3rd Joint IFIP Wireless and Mobile Networking Conference, WMNC'10, Budapest, Hungary, October 2010. As chapter 3.
3. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Distribute the Video Frame Pixels over the Streaming Video Sequence as Sub-Frames. The 4th International Conferences on Advances in Multimedia. Chamonix / Mont Blanc, France, May 2012, pp. 133-141. As chapter 4.
4. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Eliminating the Effects of Freezing Frames on User Perceptive by Using a Time Interleaving Technique. Journal of Multimedia Systems, Vol. 18, Issue 3, 2011, pp. 251-256. As chapter 5.
5. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Compressing Video Based on Region of Interest. To Appear in the EUROCON 2013 Conference Proceedings. As chapter 6.
6. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Adapting the Streaming Video on the Estimated Motion Position. Journal of Advance in Electrical and Electronic

Engineering, Special Issue on Information and Communication Technology and Services, Vol. 10, No. 4, 2012, pp. 240-245. As chapter 7.

7. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Identifying the Position of the Motion Region of Interest to be Adapted for Video Streaming. Submitted for publication. As chapter 8.

The following publications are associated with, but not included in this thesis:

8. 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.
9. Hussein Muzahim Aziz, Håkan Grahn, Lars Lundberg. Sub-Frame Crossing for Streaming Video over Wireless Network. The 7th International Conference on Wireless On-demand Network Systems and Services, WONS'10, Kranjska Gora, Slovenia, February 2010, pp. 53-56.
10. Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, Lars Lundberg. Adapting the Streaming Video Based on the Estimated Position of the Region of Interest. The 8th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS'12, Sorrento - Naples, Italy, November 2012, pp. 1-7.

Paper 9 is an earlier version of paper 3 and paper 10 is an earlier version of paper 7.

Table of Contents

Chapter Six

Chapter Seven

Chapter Eight

CHAPTER ONE

1.1 Motivation

Real-time video streaming over wireless networks often suffers from bandwidth limitations and outages, which may lead to frame losses. Video frame losses have a significant effect on the user-perceived quality. My research goal is to provide a satisfactory video quality for real-time video streaming over unreliable networks with limited bandwidth.

Real-time video streaming over unreliable networks requires special techniques that can overcome the loss of video frames and reduce the amount of data that are transmitted to the mobile device. In this thesis, we propose several techniques to overcome the network limitations and to provide a smooth video playout with a satisfactory quality to the mobile users.

1.2 Background

This chapter presents a background study of the thesis work for streaming video over wireless networks. The streaming environment study is based on three entities; the streaming server, the transmission media, and the mobile device. In the streaming server, we implement our proposed streaming techniques for the chosen videos. The chosen videos are considered as professional test videos with different characteristics and different important regions. The videos are coded with ffmpeg [60]; ffmpeg is considered one of the most popular software packages used in academia. The encoded videos are transmitted either over multiple channels or over a single channel, according to the techniques that are implemented on the streaming server. The purpose of the proposed techniques is to avoid loss of video frames over limited bandwidth channels. In case there is a missing frame on the mobile device, a reconstruction mechanism will be applied. The reconstruction mechanisms are different from one proposed technique to another; either we applied spatial or temporal interpolation depending on the proposed streaming technique.

1.2.1 Video Coding

The Advance Video Coding (AVC) for H.264 has been developed jointly by ITU-T's Video Coding Experts Group (VCEG) and ISO/IEC's Moving Picture Experts Group (MPEG) [36]. The video coding standard MPEG-4 / H.264 (AVC) removes redundancy or similarities between the neighbouring frames in the video sequence [17, 48]. There are two kinds of redundancies present in video frame sequences: spatial and temporal redundancy [15, 29, 48].

The spatial characteristics of the video scene, such as the one shown in Figure 1, that are relevant for video processing and compression are, e.g., texture variation within the scene, the number, colour and shape of objects. Temporal characteristics are, e.g., object motion, changes in illumination, and camera movement [15].

Temporal quality refers to the number of frames per second, where the motion in a scene appears smoother if more frames per second are played. The transmitted frames should be received and played according to their deadlines, otherwise the video is frozen. Spatial quality can be expressed as the number of pixels in the video frame, where the texture frame provides information about the spatial segmentation or selected area of interest in the video frame. The size of the video frame resolution can be changed to fit into different display screens. The resolution can be changed for both reduction and expansion, by removing and adding pixels in different parts of the video frame [44]. The video frame is blurred or juddered if a pixels or blocks of pixels are affected by artifacts. Both temporal and spatial quality, as shown in Figure 2, are usually determined before video encoding, preventing dynamic trade-offs during the encoding process [51].

H.264 (AVC) is the most popular standard for video coding [55], and it can provide an excellent compression ratio [13]. Video transmission over error prone environments and limited bandwidth channels will affect the video quality; therefore there is a need for video scalability. Video scalability refers to the removal of parts in the video stream in order to adapt it to various needs according to the end-user's specification and the terminal capabilities, as well as the

network's condition [14]. Scalable Video Coding (SVC) is an extension of the H.264 (AVC) standard that supports temporal and spatial scalability. Temporal scalability refers to the number of frames that can be removed from the streaming video sequence [55]. Spatial scalability refers to scalability with respect to the resolutions of the video frames [28].

Figure 1: Still image from a video scene.

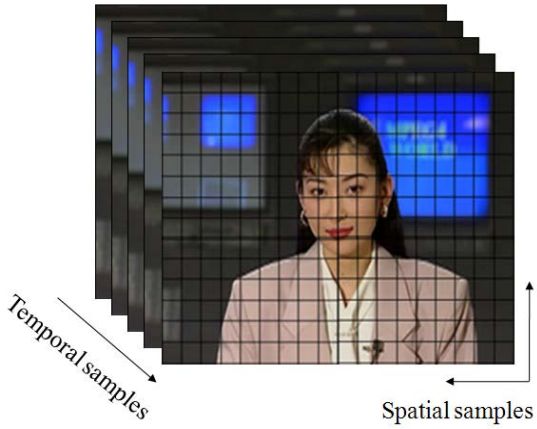


Figure 2: Spatial and temporal sampling of a video sequence.

Flexible Macroblock Ordering (FMO) is an error resilience tool that is introduced in the H.264 standard [13, 55]. FMO divides the video frames into slice groups; the maximum number of slice groups per frame is eight [24]. Each slice group could contain one or more slices, where each slice is a collection of macroblocks (MBs) [13, 24, 55]. Each MB contains a block of pixels [14]. This means that FMO makes it possible to split each frame into different parts (slice groups) with at most eight slice groups per frame.

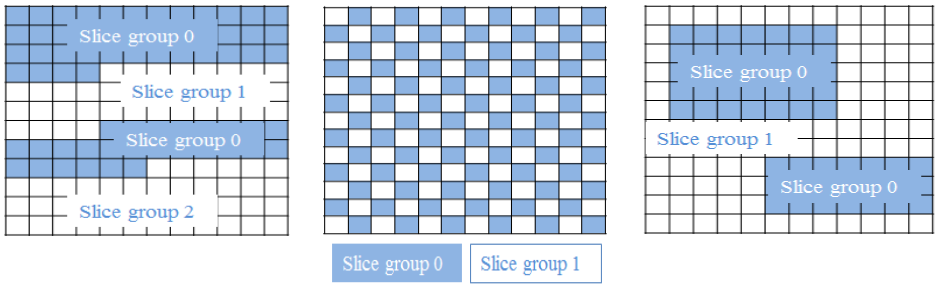
FMO consists of seven slice group types. In type 0, each slice group consists of a numbers of MBs and each MB is assign in sequential order before another slice group is started, as shown in Figure 3. In type 1 a predefined function is used to create a dispersed pattern, as the MBs are assigned to the slice groups according to the number of the slice groups. If the MBs are separated into two slice groups; it will create a chessboard pattern, the MBs can also be separated into four or more slice groups, as each slice group will not contain neighbouring MBs. Type 2 defines one or more rectangular slice groups and a background slice group within the video frame. In types 3 to 5, the MBs are separated over two different slice groups only. In type 3, the slice group will start from the centre in each frame and it will grow outward. In type 4, a certain amount of MBs will be added to the slice group from top to bottom or vice versa. Type 5, is the vertical counter part of type 4, where one slice group will be mapped from left to right or vice versa. The seventh type is used when the MBs cannot be represented by any of these six types and the MBs can be assigned to slice groups according to any user defined mapping function [13, 36, 55]. The advantage of FMO is to separate the MB within the video frame to slice groups. These slice groups can be encoded and transmitted independently from each other [36].

Scalable Video Coding (SVC) uses Flexible Macroblock Ordering (FMO) to makes the H.264 standard more flexible and adjustable for different network technologies and user requirements [28]. SVC provides a bit rate adaption for varying channel conditions and extracting the important slices with high quality [14]. SVC can handle different representations of the same video frame, by partitioning the video frame into important and less important slices [13], thus

enhancing a specific region of the frame for both spatial resolution and quality [28].

The important slices in the video frames are considered as the Region-Of-Interest (ROI). The ROI is an area within the video frame that contains visual information that is more important to the viewers than the other parts of the video frame. The other parts are called the non-Region-Of-Interest (non-ROI) [37,38].

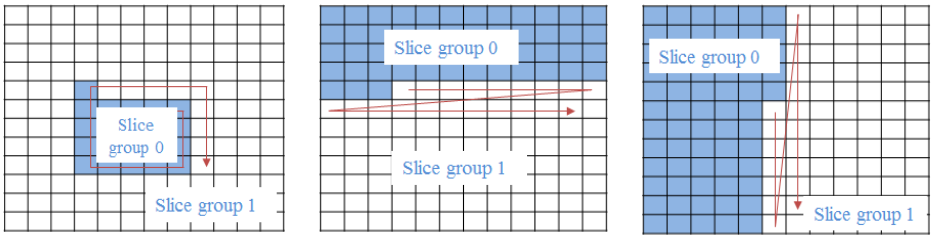
In this thesis, we have looked at techniques that are similar to the ones used in video scalability. However, we have implemented these techniques on a higher level by splitting the original videos before they are coded by H.264. The H.264 codec has become an attractive and widely used candidate for real-time video streaming [45,49], because it can achieve high compression [18].



Type 0 : Interleaved

Type 1 : Dispersed

Type 2 : Foreground and Background



Type 3 : Box - out

Type 4 : Raster scan

Type 5 : Wipe – out

Figure 3: The types of FMO in H.264/AVC standard.

1.2.2 Video Transmission

Video transmissions over wireless networks use two modes, namely downloading and streaming. In download mode, an entire video file will be downloaded from the video server and then the video file will be played. In streaming mode, all video content does not need to be downloaded before viewing. Instead the video is played when a sufficient number of video frames have been received by the mobile device [6, 7].

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 [50]. Video transmission over a dynamic channel, like mobile wireless networks, is much more difficult than over a static channel, since the bandwidth, delay, and packet loss are not known in advance [47]. Therefore, wireless networks need some effort in order to use channel resources efficiently [19].

Forward Error Correction (FEC) is a technique that is commonly used to handle packet losses over the network. FEC has become a suitable technique for video streaming due to a small transmission delay time that can offer error protection to improve the transmission reliability by adding extra redundant packets (parity packets). The principle of FEC is that the video data are packetized into k source packets with h redundant packets which become a block of n packets. The blocks of n packets will be transmitted to the receiver over the network. If there is a packet lost during the transmission, the lost packet can be successfully recovered from the redundant packets at the receiver side [5].

The availability of multiple channels for wireless communication provides an excellent opportunity for performance improvements. The term multiple channels refers to wireless technology that can use more than one radio channel [20]. An advantage of using multiple channels compared to a single channel is increased channel capacity for video transmission [4].

Multiple-Input-Multiple-Output (MIMO) technology is an effective method to reduce the fading impairments [52] and deliver very high

spectral efficiency for the transmission of information over wireless channels [25, 26]. 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 that of a single path.

Multi-path transport can improve reliability by overcoming the congestion problem often encountered in the single-path case [21]. Transmitting the video through MIMO is based on Spatial Multiplexing (SM). SM increases 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 mobile terminal via *N* antennas as well (currently most terminals have one antenna with the possibility that they could have *N* antennas in the future [41]). The transmitted data can be sent up to at *N*-times the rate of a standard terminal [32].

Multiple Descriptions Coding (MDC) is a source coding technique that generates multiple correlated bitstreams, each of which can be independently encoded and decoded [20]. Each bitstream, called a description, is transmitted through the networks and is expected to follow a different path to reach the destination. The transmission time, delay, and errors that could occur are independent and, hence, the performance of MDC is high [22, 53]. Therefore, MDC is considered as a promising technique to enhance the error resilience for video transport systems [54] by transmitting the video over multiple independent channels like MIMO, as shown in Figure 4.

Figure 4: Transmitter and receiver structure.

1.2.3 Transmission Protocol

Communication between devices in order 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 popular protocols used to communicate between network devices. TCP is connection-oriented, reliable with flow control. TCP uses an acknowledgement and retransmission scheme to guarantee the delivery of data. In contrast, UDP provides an unreliable and connectionless communication service [1, 27].

Most real-time video services employ UDP as the transport protocol [10]. Compared to TCP, UDP does not involve any retransmission mechanism, which makes it attractive to delay-sensitive applications [12, 17, 27]. Further, UDP does not perform any error recovery. Streaming video over wireless networks must often meet strict real-time requirements. This makes retransmission problematic or even impossible. Therefore, UDP is considered as a good candidate for real time video streaming [10, 17]. However, in order to be resilient against outages, UDP must often be complemented with foreword error recovery mechanisms, e.g., interleaving and interpolation recovery mechanisms, like the ones that are considered in this thesis.

In order to stream the video in real time, the Real-time Transport Protocol (RTP) must run on top of UDP to make use of RTP services. RTP will provide an end-to-end delivery service. Those services include payload type identification, sequence numbering, time stamping, and delivery monitoring. RTP does not provide any mechanism to ensure timely delivery or provide other QoS guarantees; instead it relies on lower-layer services [27].

1.2.4 Quality of Service

Quality of Service (QoS) refers to the ability to provide a satisfactory service during a communication session [40]. Consistently anticipating and meeting users' QoS needs are what distinguishes successful communication services and product providers from their competition [40]. Several parameters could affect the quality of video transmission over wireless networks, e.g.,

Compression parameters. The main issue that makes video streaming difficult is the large number of video frames that are transmitted over wireless networks. However, video streaming is compressed in a lossy manner by H.264 codec, leading to smaller representations of video data than those that are available with lossless data compression. Compression plays an important role in video streaming [6, 11, 45, 46]. The nature of the video scene, like the amount of motion, colour, contrast, frame size, and the number of frames that are transmitted per seconds, can also have an impact on the human perception of the video quality [6, 11, 15, 46].

Wireless network parameters. The main issue with real time video streaming is that it is difficult to guarantee an end-to-end QoS during the entire streaming process [49]. The current best-effort networks do not offer any QoS guarantees for video transmission over wireless networks [6]. Wireless network performance is defined as the requirements that must be guaranteed, such as bandwidth, end-to-end delay, and jitter. The network services depend on the traffic behaviour and perform according to the traffic parameters such as the peak data rate [6, 17, 47]. Further, frame losses during the transmission could have a negative effect on the quality of the video.

1.3 Objectives and Research Questions

Real-time video streaming over wireless networks has become very popular nowadays due to the wide spread use of computer laptops, and mobile devices. The transmission rate of wireless channels varies from time to time and it depends on the available bandwidth. Wireless channels are unable to guarantee the number of video frames that are transmitted to the mobile devices. The video frames could be lost, delayed, or affected by errors and become unreadable by the decoder.

The main objective of this thesis is to improve the end user's perceived quality of real-time video streaming over unreliable networks with limited bandwidth. This can be achieved by eliminating the frozen pictures and reducing the amount of video data that are streamed to the mobile device.

A number of techniques are proposed to overcome the effect of network outages and bandwidth channel limitations. Space-division techniques are based on splitting each video frame into sub-frames (slices), and transmitting these over multiple channels. Multiple channels can be used to improve fault tolerance; link recovery can also provide larger aggregate bandwidth. Interleaving is another technique that can be used to eliminate a sequence of lost frames, by reordering the frames in the streaming sequence. The size of the video can be reduced by identifying and transmitting the important region, which we call the Region-Of-Interest (ROI) and drop the less important region, which we call the non-Region-Of-Interest (non-ROI). The techniques are designed in a way that can improve the end-user's perceived quality. In this context, we address two primary research questions and these research questions are broken down into two secondary research questions each.

Primary research question one (RQ1): How can we provide a smooth video playback with satisfactory video quality to the mobile viewer in the presence of frame losses?

The research question is dealing with different aspects, and RQ1 is divided into two secondary research questions.

> **Secondary research question one (RQ1.1):** How can we improve the user's perceived real-time video streaming quality by exploiting space-division techniques on the video frames?
>
> **Secondary research question two (RQ1.2):** How can we improve the user's perceived real-time video streaming quality by exploiting time interleaving techniques on the video frames?

Primary research question two (RQ2): How can we use the Region-Of-Interest concept to reduce the amount of video data that are streamed to the mobile device over the wireless network and provide a satisfactory video quality?

The research question is dealing with different aspects, and RQ2 is divided into two secondary research questions.

Secondary research question one (RQ2.1): How can we adapt the streaming video based on identifying and extracting a fixed Region-Of-Interest?

Secondary research question two (RQ2.2): How we can adapt the streaming video based on identifying and extracting the most high motion region and use this as a Region-Of-Interest?

The research questions address the objectives of this thesis work. The proposed techniques that answered the questions with their results are presented in chapters two to eight. Each chapter discusses one of the techniques that are proposed in this thesis.

1.4 Research Methodology

The environment used to simulate the proposed techniques is Simulink [59]. Simulink is an interactive software that has become the most widely used software in academia and industry for modelling and simulation. Simulink provides features that allow us to stream the video in real-time.

Subjective and objective metrics are used to measure the quality level of the reconstructed videos that are obtained from the proposed techniques. The most common subjective metric used for evaluating video quality is the Mean Opinion Score (MOS). MOS has been recommended by the International Telecommunication Union Recommendations (ITU-R) [16]. MOS is a subjective quality metric that is obtained from a panel of human observers. It has been regarded for many years as the most reliable form of a quality measurement technique [31]. MOS is very tedious, expensive, and time consuming [57]. The subjects provide the perception of quality on a continuous linear scale which is divided into five equal regions marked with the MOS-scale adjectives described in Table 1.

The natural way to compare the quality of the videos is to evaluate it by human observers. Typically, an observer examines a set of videos in a controlled environment and assigns a numerical score to each video.

Table 1: Subjective quality rating scales [16].

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

Data are gathered from the individual viewers during the subjective experiment according to the MOS-scale. 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 MOS are obtained as the arithmetic mean of the scores as (1)

$$\mu = \frac{1}{N}\sum_{i=0}^{\mathrm{N}} \mathrm{u_i} \tag{1}$$

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

$$[\,\mu - \delta\,, \mu + \delta\,] \tag{2}$$

It is noted that the deviation term δ in (2) can be derived from the standard deviation σ and the number N of viewers that give 95% confidence interval according to ITU-R recommendations:

$$\delta = 1.96\frac{\sigma}{\sqrt{N}} \tag{3}$$

where the standard deviation σ, is defined as the square root of the variance

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

The sample size for each study in this thesis is 30 persons and the output results will be based on their observations. The experiments have taken place in the university library. The physical environment is set according to the ITU-R recommendations. Each person evaluates and judges the video based on a 5-score scale as shown in Table 1. The observers who are involved in the study should not have any professional backgrounds on video processing. The measurement method is applied to all the videos that are played for the same observer. The evaluation form should be marked by the observer each time the video is played.

The Structural Similarity (SSIM) index is an objective metric [58] used to evaluate the overall video quality level based on comparing the test video frames often compression and decompression with the full-reference frames (original frames) video according to formula (5),

$$S(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

Where *x* and *y* are the two frames to be compared, and μ_x, μ_y, are the mean of *x*, the mean of *y*, respectively. Where x_i and y_i denote the *i-th* pixels (*i*=1,2,…,*N*) of the original frame vector *x* and *y*, respectively and according (6),

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \ \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (6)$$

where σ_x , σ_y , are the standard deviation of *x*, and the standard deviation of *y*, respectively (7),

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{\frac{1}{2}}, \ \sigma_y = \left(\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2\right)^{\frac{1}{2}} \quad (7)$$

where σ_{xy} is the covariance of *x* and *y*, according to (8),

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \quad (8)$$

While, C_1 and C_2 are constants defined as $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ where K_1 and K_2 are constant experimentally determined ($K_1 = 0.01$ and $K_2 = 0.03$), as they help to improve the stability of the measure when the denominator is close to zero.

The calculation is based on pixel-by-pixel that is moved from top-left to bottom-right corner of the frame. The results that are obtained from the SSIM index value is between 0 and 1, where the value 1 is only reachable when the two frames are identical.

The experimental study is carried out for the different techniques that are proposed in this thesis, and the test videos are used with a resolution of 176 x 144 pixels. As mentioned before, the chosen videos have different characteristics.

The open source ffmpeg codec software [60] is used in this thesis to encode/decode the test videos. The reason of using the ffmpeg codec is to study the effect of the proposed technique on the video size. The compressed video sizes are compared between the original videos and to the videos that are obtained from the proposed techniques.

1.5 Thesis Contributions

In this thesis, we propose several techniques for streaming video over unreliable networks. The proposed techniques can be adapted to the current network conditions, requirements, and could possibly help to overcome the effect of bandwidth variation and outages in video streaming. The main contributions of this thesis are:

Chapter two: A novel frame splitting technique is proposed to create two sub-frames out of each frame, where one sub-frame contains the even pixels and 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 risk that two sub-frames related to the same frame are lost due to outages in the wireless channel, particularly when the two logical channels (one for even and one for odd pixels) are multiplexed on the same physical wireless channel. 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 three: An extension of the technique in chapter two is proposed, where each frame is split into four sub-frames and each sub-frame contains one fourth of the main frame pixels. The four sub-frames will be encoded by MDC using the H.264/AVC codec. The encoded sub-frames will be transmitted over multiple channels and each sub-frame represents 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 be delayed for 1.5 seconds. The reason for 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 four: An extension of the technique in chapter three where each frame is split into four sub-frames and each sub-frame will then be combined with another sub-frame from a different sequence position. The combined sub-frames will be transmitted as a single frame over a single wireless channel. In case of frames losses or frames corruption from the streaming video, there is still a possibility that one of the sub-frames will be received by the mobile device. A rate adaptation mechanism is also highlighted in this work. The server can skip up to 75% of the frame pixels (three sub-frames) and we can still be able to reconstruct the received sub-frame with acceptable quality and play it on the mobile device.

Chapter five: A novel interleaved approach is proposed to reorder the video frames before they are streamed over wireless networks. The Time Interleaving Robust Streaming (TIRS) is proposed to eliminate the frozen picture by avoiding a sequence of neighbouring frames to be lost and allow at least every second frame to be present on the mobile device. The TIRS technique will group the frames as sequences of even frames followed by odd frames and compressed by H.264 codec. The benefit of streaming the video according to TIRS is to make it possible to reduce the effects when a sequence of consecutive video frames is lost during the transmission. If a sequence of frames is lost, TIRS will

spread out the lost frames in the streaming video sequence with the ability to reconstruct the missing frames at the mobile device.

Chapter six: A new technique is proposed to reduce the amount of data by adapting the video frames that are streamed over limited channel bandwidth. The streaming server will identify and extract the high motion region (ROI) from the frames that are between reference frames and drop the less motion region (non-ROI). Four different ROIs for three different scenarios are used to study the effect of the compression size on the video streaming. When the mobile device starts receiving the streaming video, linear interpolation will be performed between reference frames to reconstruct the pixels that are outside the ROI (non-ROI).

Chapter seven: An extension of the technique in chapter six where the estimated position of the ROI are defined and extracted from the frames that are between reference frames on the streaming server. The position of the ROI will be different from one video to another, while the amount of pixels that are transmitted is fixed. Two scenarios are proposed to encode the videos using H.264. In the first scenario; the original videos are encoded with default quantization parameters (QP) for a low bit rate, while for the proposed scenario, the videos are encoded with a higher bit rate. The main idea to encode the videos in the second scenario with adaptive QP is to gain equivalent encoding size for the videos that are in the first scenario that can cope with the limitation of the bandwidth channel. Linear interpolation is applied to reconstruct the non-ROI from reference frames on the mobile side.

Chapter eight: An extension of the technique in chapter seven where the size of the ROI and the non-ROI will be different from one video to another, as each video has different characteristics and different motion levels. Therefore, the amounts of data that will be dropped are different from one video to another. Linear interpolation is applied to reconstruct the non-ROI from reference frames on the mobile side. The original video is coded with a low bit rate, while the proposed scenario is coded with a high bit rate. The observer evaluates the videos after the non-ROI has been reconstructed.

1.6 Research Validity

The results that are obtained from the system design that has been evaluated by subjective and objective metrics are quite encouraging, but there are some limitations that need to be highlighted.

Regarding the simulation study, we design our techniques in a way that allows us to stream the videos frames from the server node to the client node without considering the communication setup time between them, which we assume to be 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. Therefore, only the frame losses and outage are considered in the simulation environment, as it is one of the main focuses in this thesis.

The chosen videos for the experimental studies had different characteristics and motion levels. The number of videos are used in each experimental are different from one proposed technique to another, as four videos are used in chapters two to five, and five videos are used in chapters six and seven, while eleven videos are used in chapter eight. If more videos are considered with different variety of objects and scenes, then we could have better vision of the techniques that are proposed in this thesis.

During the subjective experimental study, some of the user panels requested to re-play the videos as the videos length is short. They felt that, they need to play the videos again for fair judgments. Miras [8] also highlight that the video sequence is not long enough to experience different kinds of impairments that could occur to the video test sequence. Therefore, the lengths of the videos are extended by repeating the same video several times to achieve the purpose and the requirements of the scenarios that are proposed in chapters 2, 3, and 5.

The number of human observers that are considered by the ITU-Recommendation is at least 15 observers for each Mean Opinion Score (MOS) study [16]. In order to improve the reliability for each experimental study, the number of observers that evaluate each proposed technique is 30.

The main measurement index used to evaluate the videos in all chapters in this thesis are the MOS metric, except the videos that are obtained from chapter three is evaluated by SSIM metric. The reason for that is to avoid repeating the same experimental evaluation to the reconstructed videos that are obtained from chapter two, therefore SSIM metric are used in chapter three to measure the quality level of the videos.

The videos that are obtained from each chapter could be measured by using two metrics to get an accurate evaluation, as an example MOS measurement as a subjective metric and SSIM index as an objective metric. The SSIM index use a full-reference frames (original frames) metric, which should be available to evaluate the similarity between the full-reference frames and the test videos frames.

1.7 Related Work

Several methods and techniques have been proposed to overcome the effect of error prone networks on streaming video.

Apostolopoulos [23] suggests that it can be beneficial to transmit different amounts of traffic on different channels, by dividing the video frames into even and odd frames that are encoded by Multiple Description Coding (MDC), and transmitted over two different channels. Transmitting the video over two channels minimizes the risk that both streams are lost or corrupted. 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 handling for that stream. The use of two channels is to provide error concealment and state recovery of the corrupted stream from the previous and future frames in the stream that was successfully transmitted.

Tesanovic et al. [30], proposed a scheme for video transmission for efficient and robust video streaming over wireless channels, through a combination of multiple channels and Multiple Description Coding (MDC) technology. The Spatial Multiplexing (SM) is used to achieve the throughput with no requirements for additional spectrum. SM relies

on transmitting independent data stream for each transmit antenna. The MDC will sample the video into two descriptions, where the two descriptions are independently coded at the encoder with the same bit rate. The video packets are split into odd and even packets and transmitted over two channels, where the packets will be reassembled at the receiver side. MDC can exploit the interactions between descriptions when losses occur on multiple wireless channels for reliably recover to the video stream. Two independent channels are used to ensure that at least one of the two descriptions is received. This can greatly improve the decoded video quality in multiple channels environments when one of the channels has failed.

Claypool and Zhu [33] proposed a different direction for avoiding the effects of channel error and frame losses during the transmission. They present an interleaving algorithm on MPEG encoded video frames. Interleaving represents a good technique for wireless networks which spreads the losses over the video stream. Interleaving of the sequence of frames is applied before the MPEG encoder. The authors suggest an interleaving distance algorithm for two and five, where two is chosen for short interleaving distance and five for long interleaving distance. The interleaving distance determines how long time consecutive frames are spread out in the video stream.

Hannuksela et al. [34] proposed a region based coding method, called sub-picture coding, which is based on rectangular shapes. The rectangular sub-picture refers to the foreground sub-picture which is the interesting region within the video frame. The arbitrarily shaped foreground sub-picture is used to define the ROI. The background sub-picture which consists of the region that does not fall into foreground sub-picture. The ROI coding and Unequal Error Protection (UEP), are two important tools are used in video streaming system to improve the visual quality of the video. ROI coding refers to techniques that allocate more bits to the ROI than the background region. UEP refers to techniques that protect a part of the transmitted bit-stream more than the rest of the stream based on portions the bit-stream in different importance. Coding the ROI will take place before the background region, as two different quantization parameters (QP) values are assigned for the picture. High QP are assign to the ROI and low QP are

assign to the leftover region (background region) with a large different of QP between the two regions. The ROI for the bit allocation method is proposed to reduce the boundary effect between the ROI and the leftover region.

Wong and Kwok [3], proposed a ROI-based channel adaptive scheme for real-time video streaming. They allocate more resource to the focus area, so that the perceived quality would be better than the area that is not in focus. The video frames will be split into two regions, the Region-Of-Interest (ROI) with high priority, and the Remainder Region (RM) with low priority. The two regions will be separately encoded by using independent video encoders. The ROI will be transmitted first. When the ROI has been transmitted successfully, the remaining available bandwidth will be allocated to the RM. A feedback message will be send to the encoders of the two regions to adjust the encoder parameters so that the network resources can be used more efficiently.

Liang et al. [56], present a framework for a content-adaptive background skipping scheme for ROI video coding. The macroblock in the video frame are classified into ROI macroblocks and non-ROI macroblocks. The content information of the video frame includes foreground motion activity, and background motion activity. The accumulated distortion due to background skipping has also been taken into consideration when the skipping decision is made. When a high motion is detected in the foreground, a decision is made to skip the background and reallocate more bits to encode the ROI for high quality. If the background contained high motion or the accumulated distortion due to background skipping is high, then the decision is not to do a background skipping.

Panyavaraporn and Cajote [24], proposed a Flexible Macroblock Ordering (FMO) slice groups generation method based on the ROI for H.264/AVC video streaming over wireless network. The method is used to generate explicit FMO slice group maps based on the combination of the ROI with spatial and temporal information. The ROI is determined by selecting the boundary of the objects in the video frame. The video frame is segmented into foreground and background, the foreground used spatial information and the background used

temporal information. The method used a number of coded bits of a macroblock (MB) that is considered as spatial information, while a distortion measure from the concealment errors is considered as temporal information which is an indicator of the important MB. A one-pass encoded scheme based on the ROI is used as feedback information to the encoder that is passed from the previous frame. The MBs are assigned to the slice groups at the encoder based on the information from previous frame.

1.8 Conclusion

Real-time video streaming is particularly sensitive to delay, frame losses and frames dropped. This is due to the variation and limitations of the channel bandwidth, which could have a negative effect on the end user's perceived quality. Every video frame must arrive to the mobile device before its playout time, with enough time to decode and display the contents of the frame. If the video frame does not arrive according to its deadline, the playout process will pause and the frame is effectively lost. The main contribution of this thesis is to improve the end user's perceived quality for real-time video streaming. Therefore, two primary research questions are addressed in this thesis. The first primary research question (RQ1) is to eliminate the frozen pictures on the mobile screen. The secondary primary research question (RQ2) is to reduce the amount of video data that are transmitted to the mobile device.

The first primary research question (RQ1) is divided into two secondary research questions, where these questions are discussed in chapter two (paper 1), chapter three (paper 2), chapter four (paper 3), and chapter five (paper 4).

The first secondary research question (RQ1.1) has been discussed in chapters two (paper 1), chapter three (paper 2), and chapter four (paper 3), where the video frames are split into sub-frames.

In chapter two (paper 1), the frames from the video sequence are split into two sub-frames, where one sub-frame contains the even pixels and another contains the odd pixels. The two sub-frames are transmitted over two independent channels with a delay time between

them. The delay time is set between the channels, is to avoid loss of sub-frames (slices) that belong to the same video frame. Missing sub-frames from any channel will be reconstructed from the received sub-frame. In chapter three (paper 2), which is an extension of chapter two (paper 1), 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 is transmitted over independent channels with a delay time between them. The missing of one, two or even three sub-frames will be reconstructed from the available sub-frames on the mobile device.

In chapter four (paper 3), which is an extension of chapter three (paper 2), a sub-frame crossing mechanism is considered based on a frame splitting mechanism. Each video frame is split into four 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 and it will be reconstructed. A rate adaptation mechanism is also discussed in this work. We show that the streaming server can skip up to 75% of the frame pixels (three sub-frames) and we are still able to reconstruct the original frame and play it on the mobile screen.

The reconstruction mechanism is used in RQ1.1 is spatial interpolation. The missing sub-frames can be rebuilt from the received sub-frames by taking the average of the neighbouring pixels to replace the pixels that are related to the missing sub-frames.

The second secondary research question (RQ1.2) is discussed in chapter five (paper 4), where the Time Interleaving Robust Streaming (TIRS) technique is proposed. The TIRS technique is proposed to eliminate the frozen pictures for relatively short outage. This is done by avoiding the sequence of neighbouring frames to be lost and make sure that for every lost frame at least two neighbour frames are present on the mobile device. TIRS will reorder the video frames as a group of even followed by odd frames. TIRS can be implemented for different interleaving times to distribute losses of frames on different positions in the streaming sequence. The interleaved videos are compressed by the H.264 codec, to study the effect of TIRS on the video file size.

The reconstruction mechanism is used in RQ1.2 is temporal interpolation. The missing frames will be rebuilt from neighbouring frames by using linear interpolation.

The second primary research question (RQ2) is divided into two secondary research questions. The first secondary research question (RQ2.1) is discussed in chapters six (paper 5), and seven (paper 6). The second secondary research question (RQ2.2) is discussed in chapter eight (paper 7). These questions considered techniques to identify the important part and less important part within the video frame. The important part within the video frame, is called the Region-Of-Interest (ROI), while the less important part, is called the non-Region-Of-Interest (non-ROI).

In chapter six (paper 5), the streaming video frames will be adapted by identifying and extracting a fixed region with high motion and spend less capacity on the region with less motion. Four different types of Region-Of-Interest (ROI) are proposed; for three different scenarios. In the first scenario the reference frames are every 3rd frame, e.g., 0, 3, 6, 9,..., in the second scenario the reference frames are every 4th frame, e.g., 0, 4, 8, 12,..., in the third scenario the reference frames are every 5th frame, e.g., 0, 5, 10, 15,... . The scenarios are designed to place the reference frames (full original frames) in different sequence positions in the streaming video. The ROI can be extracted from the frames that are between reference frames and drop the non-ROI. The four ROIs for the three scenarios are compressed by H.264 codec to study the effect of the proposed techniques on the streaming size. The reconstruction mechanism is performed in the mobile devices to rebuild the frame that belongs to the ROI and from the reference frames.

In chapter seven (paper 6), which is the extension of chapter six (paper 5), the reference frames position, are set as every fifth frame in the streaming video. The ROI will be extracted from the frames that are between reference frames. Two scenarios are discussed in this chapter and are coded by H.264, where the original video in first scenario will be coded with a low bit rate, while in the second scenario (the proposed scenario) it will be coded with a high bit rate. The quantization parameters to the second scenario will be adaptive to gain and

equivalent encoding size to the first scenario to cope with the limitation of the bandwidth channel.

In chapter eight (paper 7), we discussed the motion level within the video frames. The motion level in the video frames will be different from one video to another; therefore the size of the ROI and non-ROI will be different, as each video had different characteristics. The ROI will be extracted from the frames that are between reference frames and it will be streamed to the mobile devices. The proposed scenarios are coded by H.264 with adaptive quantization parameters to study the effect of the proposed technique on the streaming size. The ROI will be reconstructed from the frames that are between reference frames to return the frame to it is original shape.

The reconstruction mechanism is used in RQ2 is linear interpolation. Linear interpolation is applied to rebuild the pixels that are related to the ROI and from the reference frames.

The quality level for the reconstructed videos that are obtained from chapter two, three, five, six, seven, and eight; are evaluated by the Mean Opinion Score (MOS) measurement. The videos that are obtained from chapter four are evaluated by using the Structural Similarity (SSIM) index.

The results that are obtained from chapter two, three, four, and five, is significantly improves the video smoothness on the mobile device in presence of frame losses. The results that are obtained from chapter six, seven, and eight, is significantly reduce the amount of data that are transmitted to the mobile device with an acceptable video quality that can be provided to the mobile viewers.

1.9 Future Work

The thesis proposes several techniques to overcome the frozen frames on the mobile screen and reduce the amount of video data that are transmitted to the mobile devices. The proposed techniques are applied successfully to improve the smoothness of the videos that are transmitted over unreliable networks with a satisfactory video quality. There are still many open questions for future research work that are related to video streaming over wireless networks.

Avidan and Shamir [44], proposed seam-carving method to resize the frame by adjusting the frame size by adding or removing seam. The seam is defined as the path of pixels from top-to-bottom of the frame or from left-to-right of the frame, containing exactly one pixel in each row or column respectively. Seam-carving method is used to resize the frame by gracefully carving-out or inserting pixels in different parts of the frame. The energy function is used to identify the high energy pixels and low energy pixels in the frame. For frame reduction, the removing seams of a connecting path of the low energy pixels that are crossing the frame from top-to-bottom, or from left-to-right. For frame expansion, the seam insertion to the frame should be added in the way that been balanced with the original frame content. The extension of seam-carving method to resize the video frames is to treat each frame independently [35].

The work can be extended by not treating each frame independently rather than as a group of frames that belongs to the same scene. The video can be split according to the scenes; where the video could contain one or more scenes, where each scene could have different number of frames where the frames in each scene have their own characteristics.

The energy function can be used to identify the high and low energy pixels to the key-frame in each scene, where the key-frame is the first frame in each scene. The time the high and low energy is been identified in the key-frame, the seam with low energy pixels will be removed. The removal of the seam position in the key-frame will be applied to the same location to the frames that are belonging to the same scene. The expansion of the video frames is to identify and to add seam of pixels to the key-frame and applied to the same location to the frames that are belonging to the same scene.

The idea for using seam-carving method is to reduce the amount of data that are streamed to the mobile device by removing low energy seam. Since the mobile devices could have different screen resolution, enlarge the video frame will be depend on the additional pixels that are required to be added to the frames to fit the mobile screen resolution.

Removing and enlarging the size of the video based on the seam-carving method could be a promising direction in the field of video streaming over a limited bandwidth. The size of the video frames can be reduced in the server side and enlarge it in the mobile side, according to the mobile screen requirements.

References

[1] A. Löfgren, L. Lodesten, S. Sjöholm, and 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.

[2] A. Alexiou, C. Bouras, and V. Igglesis, "A Decision Feedback Scheme for Multimedia Transmission Over 3G Mobile Networks," The 2nd IEEE and IFIP International Conference on Wireless and Optical Communications Networks, WOCN'05, pp. 357-361, 2005.

[3] A. C.-W. Wong, and Y.-K. Kwok, "On a Region - of - Interest Based Approach to Robust Wireless Video Transmission," In the Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks, ISPAN'04, pp. 1-6, 2004.

[4] 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.

[5] C-H Lin, C-K Shieh, N. K. Chilamkurti, C-H Ke, and W-S Hwang "A RED-FEC Mechanism for Video Transmission over WLAN," IEEE Transactions on Broadcasting, Vol. 54, No 3, pp. 517-524, 2008.

[6] 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.

[7] 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.

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

[9] D. Grois, E. Kaminsky, and O. Hadar, “ROI Adaptive Scalable Video Coding for Limited Bandwidth Wireless Networks,” In the Proceedings of the 3rd IFIP Wireless Days Conference, pp.1-5, 2010.

[10] 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.

[11] 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.

[12] 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.

[13] H. Chen, Z. Han, R. Hu, and R. Ruan, “Adaptive FMO Selection Strategy for Error Resilient H.264 Coding,” In Proceedings of the International Conference on Audio, Language and Image Processing, ICALIP’08, pp. 868-872, 2008.

[14] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 9, pp. 1103-1120, 2007.

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

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

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

[18] J. Zhang, F. Liu, H. Shao, and 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] 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.

[24] J. Panyavaraporn, and R. D. Cajote, "Flexible Macroblock Ordering Based on Region of Interest for H.264/AVC Wireless Video Transmission," In Proceedings of the 19th International Conference on Systems, Signals and Image Processing (IWSSIP'12), pp. 384-387, 2012.

[25] 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.

[26] 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.

[27] 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.

[28] L. Ferreira, L. Cruz, A. P. Amado, "H.264/AVC ROI Encoding with Spatial Scalability," 2008. http://hdl.handle.net/10400.8/85.

[29] 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.

[30] 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.

[31] M. Martinez-Rach, O.López, P.Piñol, M.P. Malumbres, J. Oliver, and Carlos T. Calafate, "Quality Assessment Metrics vs. PSNR under Packet Loss Scenarios in MANET Wireless Networks," International Workshop on Mobile Video, MV '07, 2007.

[32] 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.

[33] 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.

[34] M. M. Hannuksela, Y-K. Wang, and M. Gabbouj "Sub-Picture: ROI Coding and Unequal Error Protection," In Proceedings of the International Conference on Image Processing, ICIP'02, pp. 537- 540, 2002.

[35] M. Rubinstein, A. Shamir, and S. Avidan "Improved Seam Carving for Video Retargeting," ACM Transaction on Graphics, Vol. 27, Issue 3, 2008.

[36] P. Lambert, W. D. Neve, Y. Dhondt, and R. Van de Walle, "Flexible Macroblock Ordering in H.264/AVC," Journal of Visual Communication and Image Representation. Vol. 17, Issue 2, pp. 358-375, 2006.

[37] P. Lambert, and R. Van de Walle, "Real-Time Interactive Regions of Interest in H.264/AVC," Journal of Real-Time Image Processing, Vol. 4, Issue 1,pp. 67-77, 2009.

[38] P. Lambert, D. D. Schrijver, D. V. Deursen, W. D. Neve, Y. D., and R. Van de Walle, "A Real-Time Adaptation Framework for Exploiting ROI Scalability in H.264/AVC," Lecture Notes in Computer Science-Advanced Concepts for Intelligent Vision Systems, pp. 442-453, 2006.

[39] R. D. Cajote, S. Aramvith, and Y. Miyanaga, "FMO-based H.264 Frame Layer Rate Control for Low Bit Rate Video Transmission,"

EURASIP Journal on Advances in Signal Processing, 2011: 63, pp. 1-11, 2011.

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

[41] 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.

[42] S.-H. Yang, C.-L. Chu, and C.-W. Chang, “An H.264/AVC Error Concealment Technique Enhanced by Depth Correlation,” In Proceedings of the International Multi Conference of Engineers and Computer Scientists, IMECS’12, 2012.

[43] S. Sanz-Rodriguez, and F. Diaz-de-Maria, “RBF-Based QP Estimation Model for VBR Control in H.264/SVC,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 21, No. 9, pp. 1263-1277, 2011.

[44] S. Avidan, and A. Shamir “Seam Carving for Content-Aware Image Resizing,” ACM Transaction on Graphics, Vol. 26, Issue 3, 2007.

[45] 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.

[46] 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.

[47] V. Vassiliou, A. Pavlos, G. Iraklis, and 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.

[48] 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.

[49] 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.

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

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

[52] 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.

[53] 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.

[54] 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.

[55] Y. Dhondt, P. Lambert, S. Notebaert, and R. Van de Walle, "Flexible Macroblock Ordering as a Content Adaptation Tool in H.264/AVC," In Proceedings for Multimedia Systems and Applications, 2005.

[56] Y. Liang, H. Wang, and K. El-Maleh, "Design and Implementation of Content-Adaptive Background Skipping for Wireless Video," In Proceedings of the International Symposium on Circuits and Systems, ISCAS'06, pp. 2865- 2868, 2006.

[57] 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.

[58] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, Vol. 13, Issue 4, pp. 600-612, 2004.

[59] www.mathworks.com.

[60] www.ffmpeg.org.

CHAPTER TWO

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 improve the perceived video quality.

Keywords

Streaming Video, Mobile Device, Frame Splitting, Multichannel, Dropping Rate, Mean Opinion Score

2.1 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 [12, 13]. It is 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 it is full frame shape. Our subjective test shows that the proposed techniques could be useful to provide a satisfactory quality to the mobile viewer.

2.2 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, 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 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 is transmitted over two channels.

2.3 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, 18], where multichannel provides information to clients via multiple channels.

Our proposed technique is to avoid the freezing picture on 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 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 Figure 1.

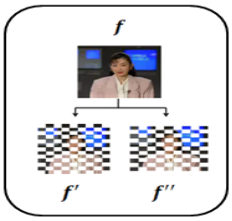


Figure 1: Frame split.

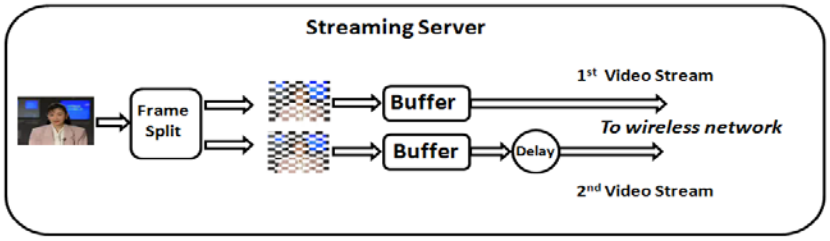


Figure 2: Streaming split sub - frames over two wireless channels.

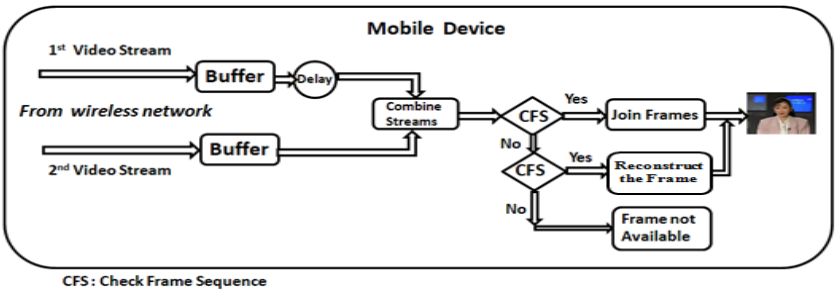


Figure 3: Receiving sub-frames streams of video on the mobile device.

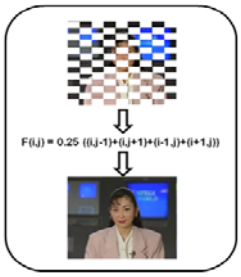


Figure 4: Reconstruct frame.

The sub-frames 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 $f' = \{ f_1', f_2', \dots f_{n-1}' \}$ will transmitted over the first stream, while, $f'' = \{ f_1'', f_2'', \dots f_{n-1}'' \}$, will transmitted over the second stream.

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 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

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

After both sub-frames 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 it is 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 is 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.

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, 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, 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%.

2.4.2 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. The user observed two scenarios, the first scenario, the observer evaluate the normal video transmission over wireless networks with the effect of three different loads, and the second scenario (the 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.

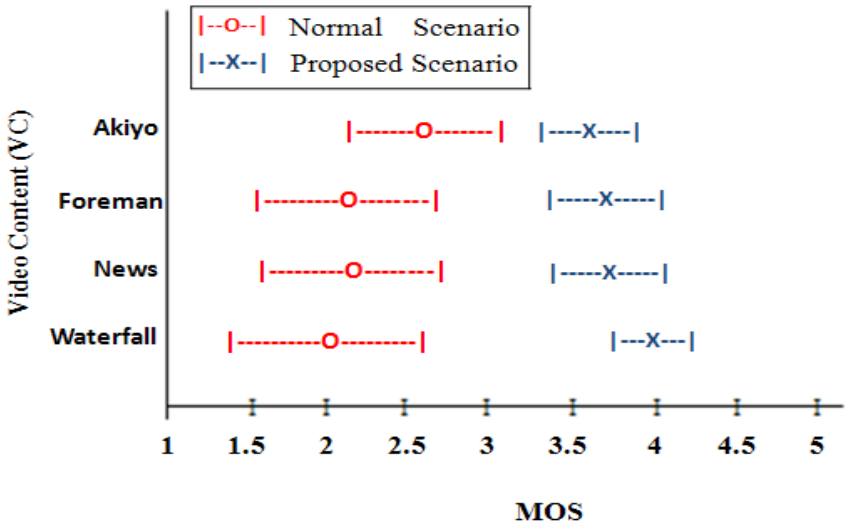
Staff and students evaluated the video quality after each sequence using a five grade MOS scale (1-bad, 2-poor, 3-fair, 4-good, 5-excellent) in a prepared form. The amount of data gathered from the subjective experiments with respect to the opinion scores that were given by the individual viewers. A 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].

2.5 Experiment Test

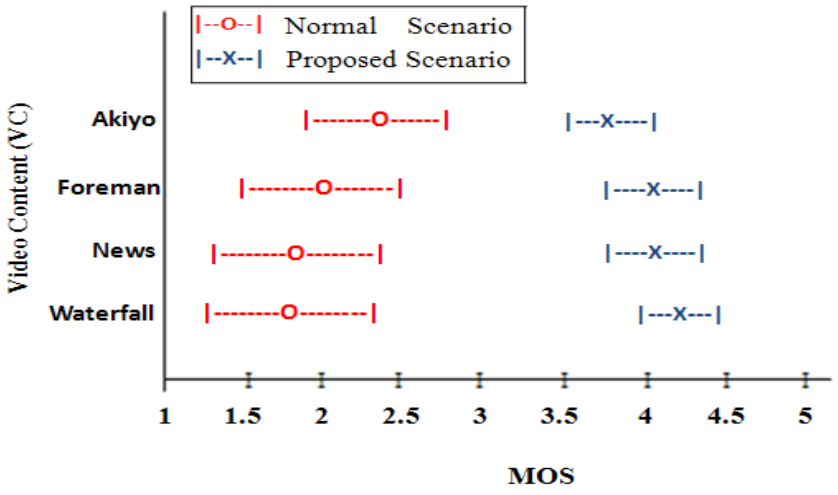
The quality of video is subjected to the personal opinion; this means that the quality of service improvements for video transmission has the only goal to satisfy the average human watching the contents of the video. The MOS is obtained through human evaluation tests, where 30 of staff and students are 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.

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 from that the video present to the viewers resulted in a wide range of perceptual quality ratings indeed for both experiments.

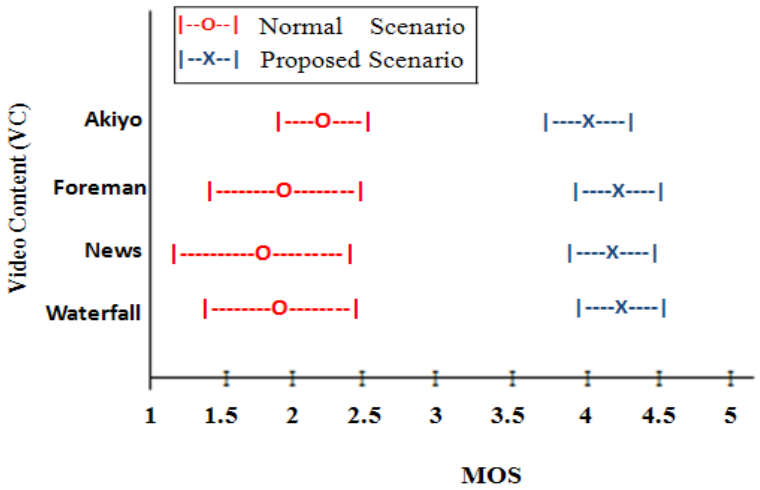
After we analysis their score we feel that our proposed scenario could be a satisfactory technique to eliminate the freezing frames when streaming videos over unreliable network.



a. Light dropped rate



b. Medium dropped rate



c. High dropped rate

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

2.6 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.

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., and 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., and 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., and Ghanbari M. 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 THREE

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 on their screen. In this work, we propose a technique to eliminate the frozen video and provide a satisfactory quality to the mobile viewers 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) measurement. 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, Mean Opinion Score

3.1 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, 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 have a major impact on the end user experience. This is particularly true in a cellular network environment where the channels condition are 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 can carry the input frames [8, 9, 10]. The objective of MDC is that, if all bitstreams have been received correctly, a high signal quality can be reconstructed. 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 on 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 on the mobile device to recreate the missing sub-frames and return it to it is 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.

3.2 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 that 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.

3.3 The Proposed Technique

Mobile real time applications like video streaming suffer from high loss rates over wireless networks [10], the result from 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 into 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 MIMO architecture.

3.3.1 Encoding the Sub-Frames

The input video frames are split into four sub-frames, where each sub-frame will represent it is 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.

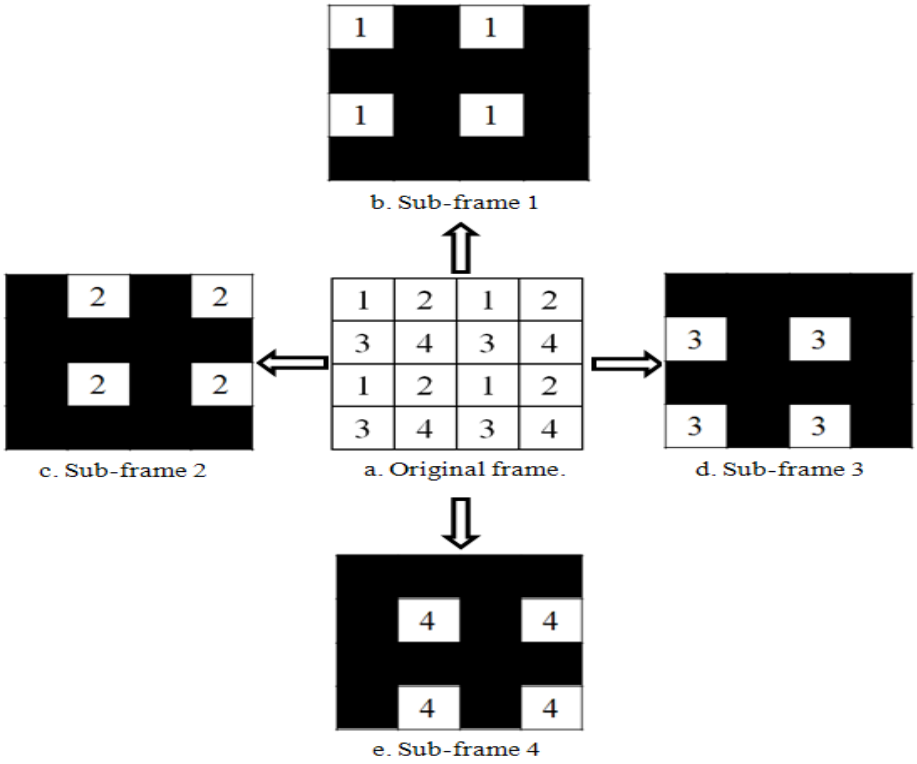


Figure 1: Frame splitting.

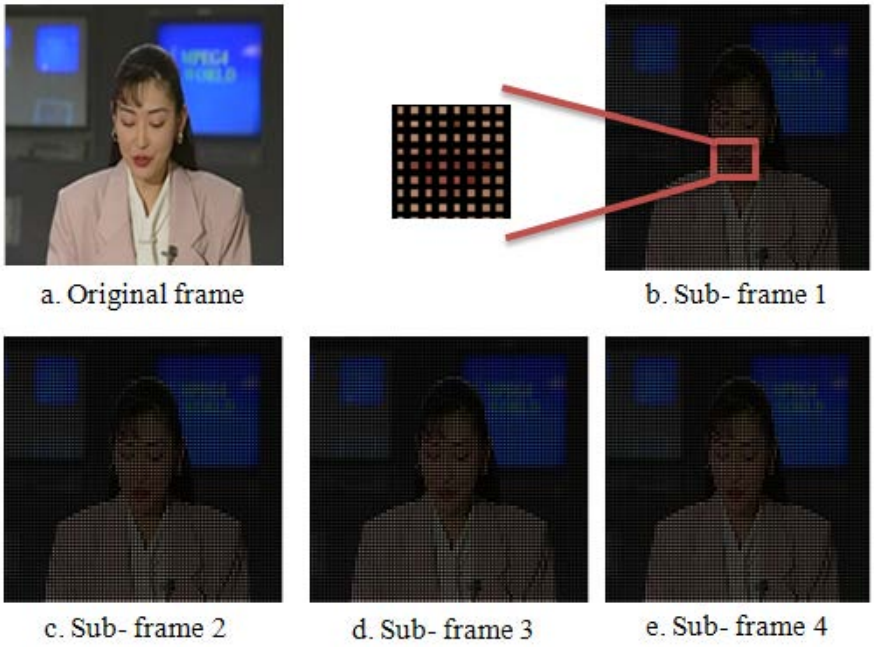


Figure 2: Snapshot of Akiyo frame splitting.

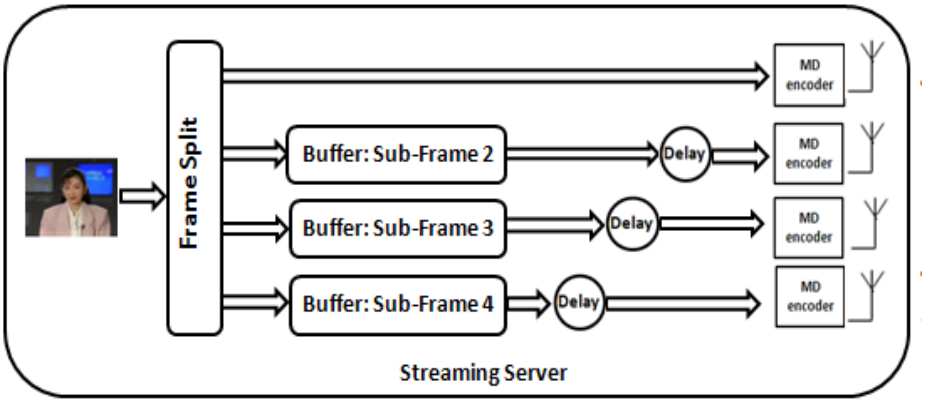


Figure 3: Streaming the sub-frames over multichannels.

3.3.2 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.

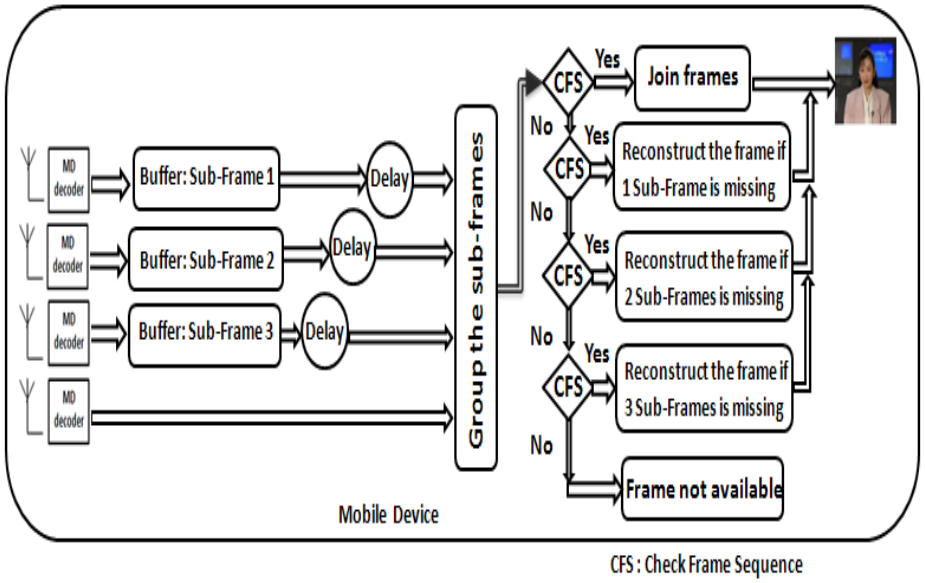


Figure 4: Receiving the sub-frames of the video on the mobile device.

The CFS and the reconstruction mechanism are used to identify the missing sub-frames and build the frame to it is 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 it is 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 it is 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 it is 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 it is normal shape.

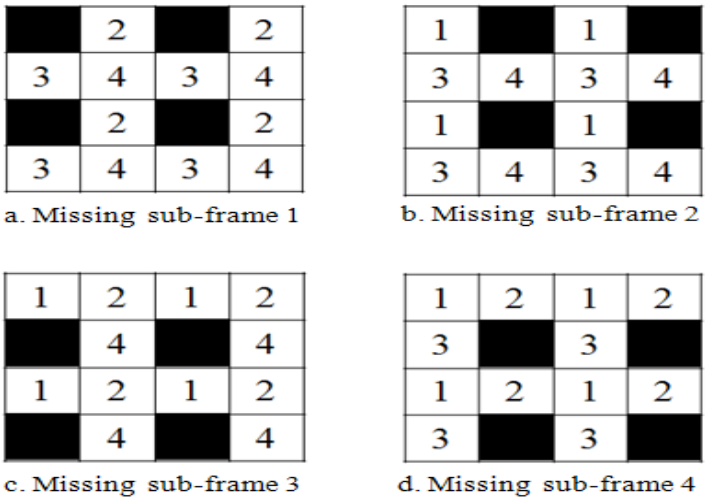


Figure 5: The possibility of missing one sub-frame.

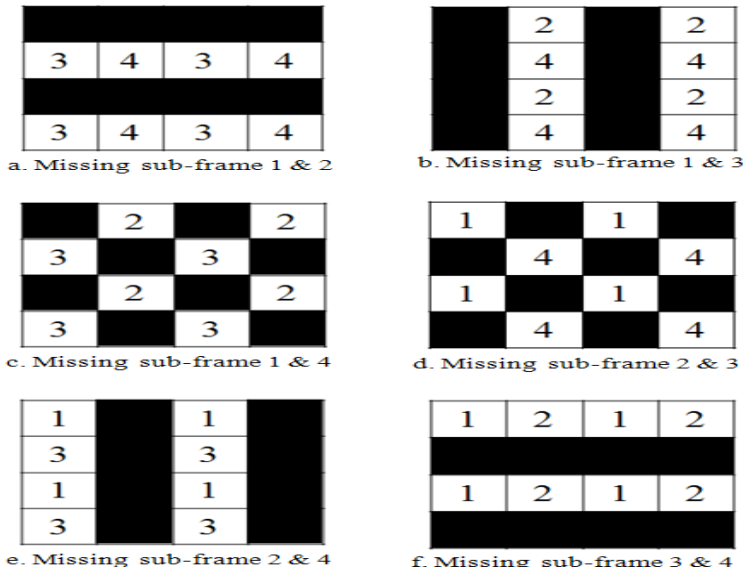


Figure 6: The possibility of missing two sub-frames.

3.4 Subjective Viewing Test

3.4.1 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 measurement 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].

3.4.2 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.

3.5 Experiment Results

The experienced quality of video is subject to the personal opinion; where the Quality of Service (QoS) improvement 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 in [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.

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 7: Snapshot of Akiyo video frame number 140.

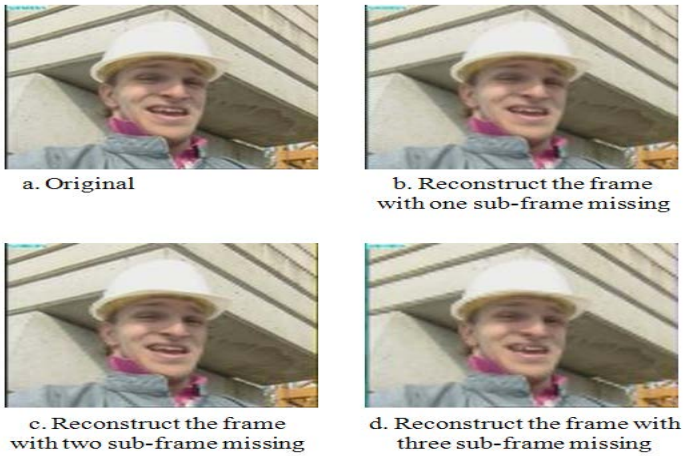


Figure 8: Snapshot of Foreman video frame number 140.

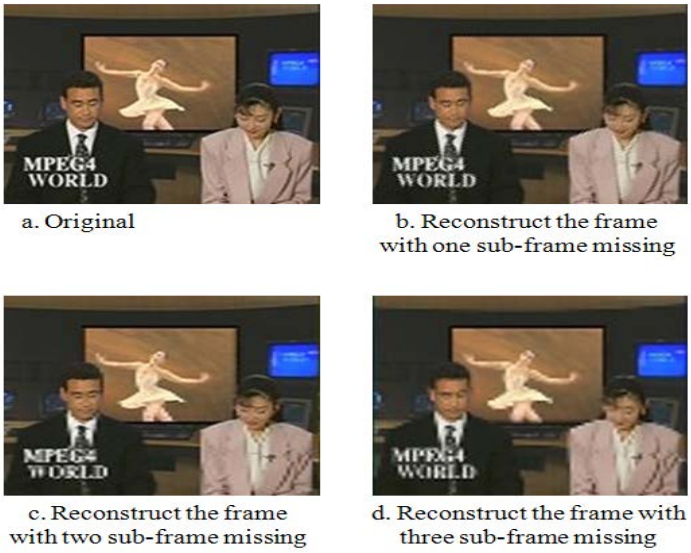


Figure 9: Snapshot of News video frame number 140.

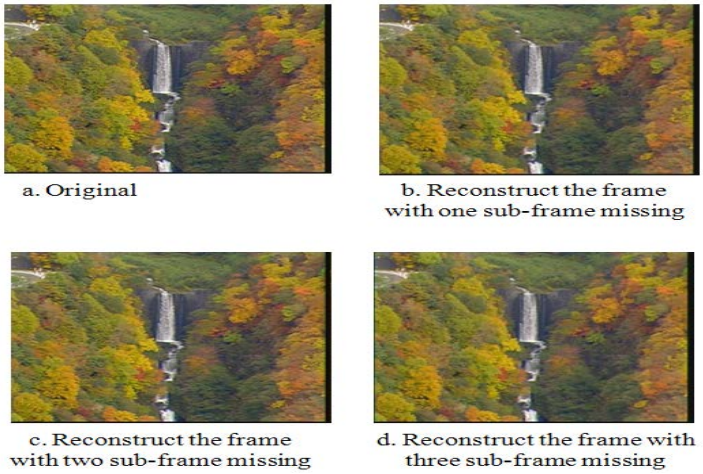
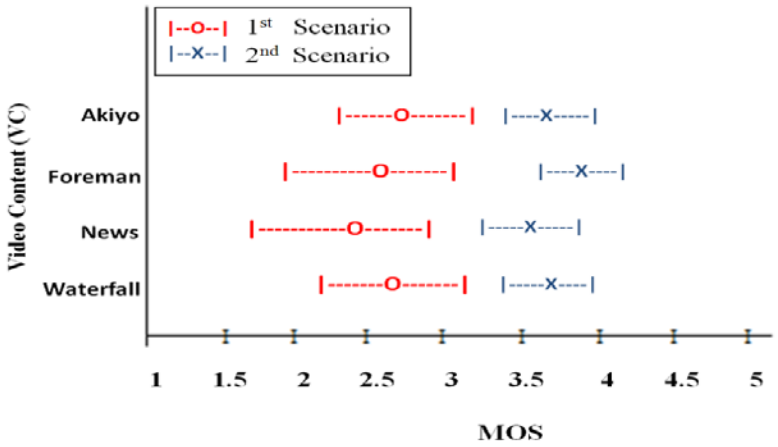
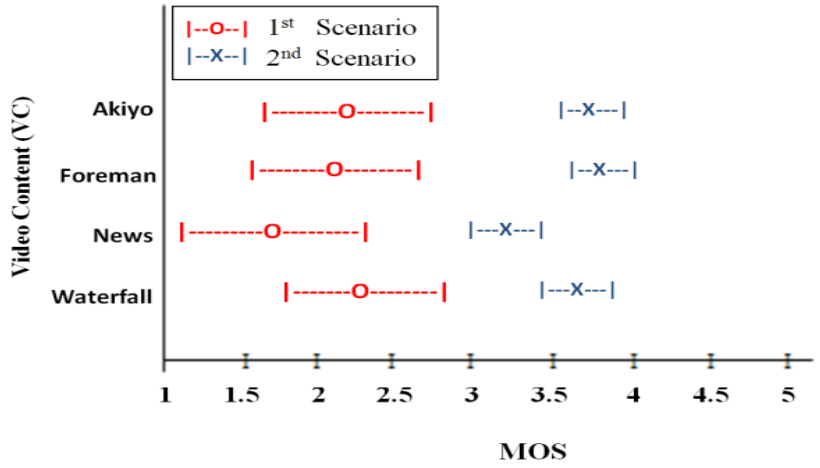


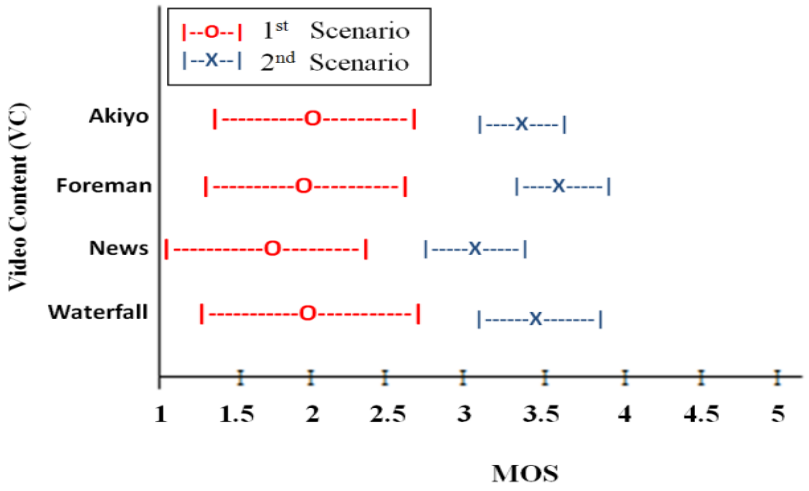
Figure 10: Snapshot of Waterfall video frame number 140.



a. Light drop rate

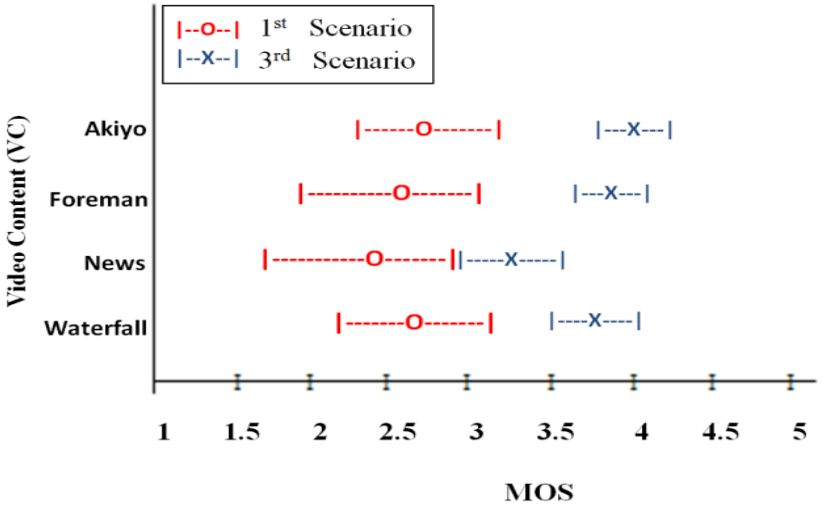


b. Medium drop rate

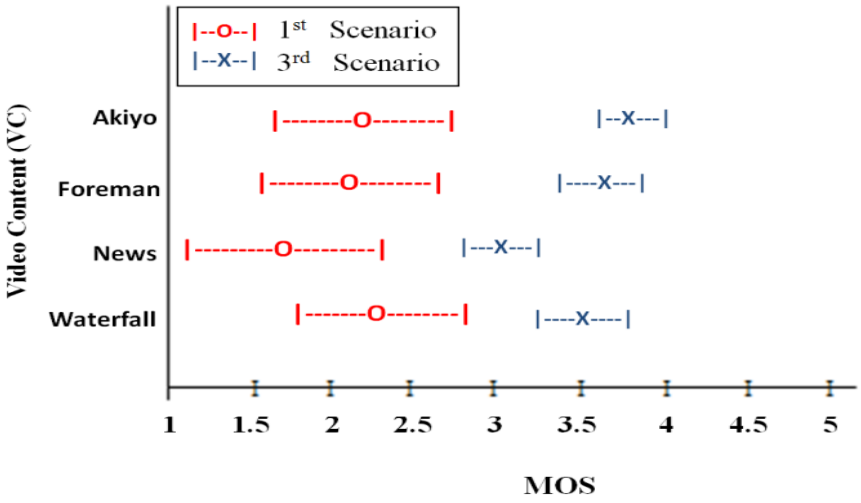


c. High drop rate

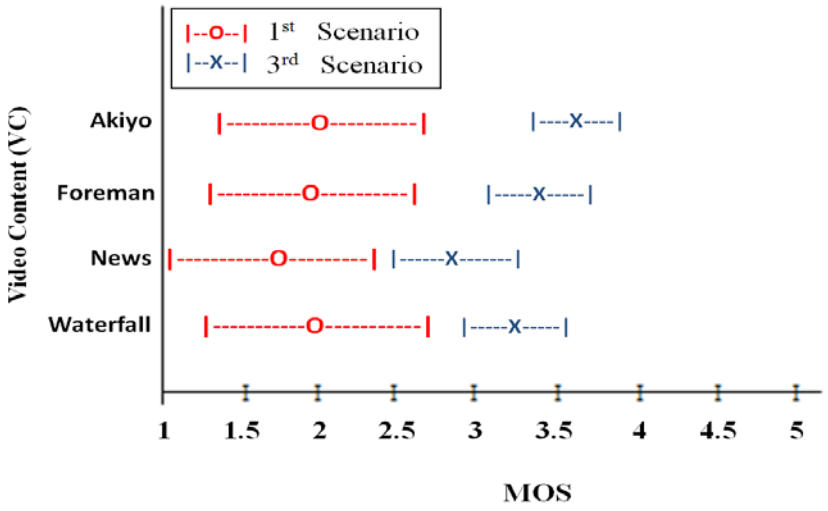
Figure 11: The MOS comparison between scenario 1 and 2, and for different video contents under different frame dropping rates.



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 from 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.

3.6 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 on the mobile device, and a reconstruction mechanism is applied to the available sub-frames to return the frame to it is 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.

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, and 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, and S. Lin, "Multiple descriptions coding for video delivery", In Proc. of the IEEE Journal, 93(1): 57-70, 2005.

[11] L. V. Andrea, A. Borneo, F. Marco, and 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, and 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, and 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 FOUR

Distribute the Video Frame Pixels over the Streaming Video Sequence as Sub-Frames

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

Abstract

Real-time video streaming over wireless channel has become an important issue due to the limited bandwidth that is unable to handle the flow of information of the video frames. The characteristics of wireless networks in terms of the available bandwidth, frame delay, and frame losses cannot be known in advance. As the effect of that, the user may notice a frozen picture on the mobile screen. In this work, we propose a technique to prevent freezing frames on the mobile devices based on spatial and temporal locality for the video stream, by splitting the video frame into four sub-frames and combining them with another sub-frames from different sequence positions in the streaming video. In case of frames losses, there is still a possibility that one fourth (one sub-frame) of the frame will be received by the mobile device. The received sub-frames will be reconstructed based on the surrounding pixels. The rate adaptation mechanism will be also highlighted in this work, by skipping sub-frames from the video frames. We show that the server can skip up to 75% of the frame's pixels and the receiving pixels (sub-frames) can be reconstructed to acceptable quality on the mobile device.

Keywords

Streaming Video, Wireless Network, Frame Splitting, Sub-Frame Crossing, Rate Adaptation

4.1 Introduction

Nowadays mobile cellular networks provide different type of services and freedoms to the mobile users anywhere and at any time, while the mobile users on the move. Streaming services become an important application to the mobile user, while 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][5][14].

The unpredictable nature of wireless networks in terms of bandwidth, and loss variation, remains one of the most significant challenges in video communications [9]. In this context, video streaming needs to implement an adaptive techniques in terms of transmission rates in order to cope with the erroneous and time variant conditions of the wireless network [9][10].

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 the streaming video should be maintained according to the networks bandwidth [2][6][11].

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 [15]. To stream video, it is desirable to adjust the transmission rate according to the perceived congestion level in wireless networks, to maintain the suitable loss level and fairly shared bandwidth with other connections. Furthermore, it is favourable for the streaming video to be aware of the transmission level in order to obtain good streaming quality by appropriate error protection.

In this paper, we proposed a sub-frame crossing technique based on frames splitting. The video frame will be split into four sub-frames, and combine the sub-frame with another sub-frame from different sequence position and from different spatial data in the streaming video. The crossing frames in the streaming video will carry pixels from four different frames that belong to four different positions and will transmit

over a single wireless channel. In case of sequence of frames losses or frames corruption from the streaming video, the losses of the sub-frames will be distributed on the streaming video and there is still a possibility that one of the fourth sub-frames will be received by the mobile device, while the missing sub-frames from the frames will be reconstructed based on the surrounding pixels.

4.2 Background and Related Work

Various techniques are proposed by many researchers for video frame slicing and reconstruction. The proposed techniques are based on H.264/AVC standard tools [20], where the Flexible Macroblock Ordering (FMO) slicing type dispersed to split the video frames and streaming them over the networks, while adaptive the slices is needed to send the highest priority information.

Huang [13] proposed a scheme for Adaptive Region of Interest (AROI) extraction and adaptation by integrating the visual attention model in the human visual system. The scheme are applied to the Region of Interest (ROI) based on video coding for adaptation and delivery, by embedding the anchor point of focusing Macroblock (MB) in each key frame and motion vectors in other frames in the coded video stream or the sequence parameter set in the Scalable Video Coding (SVC). The error resilience tool FMO can be used to define certain of ROI in SVC, while the slice groups can be used to constitute a number of columns covering the frame by some elaborated tiled partitions in order to meet the mobile terminals with different resolutions.

Wang and Tu [16] introduce an adapter FMO type, which classifies the MBs into important and unimportant slices. The important slice involves the details of the frames which represent the important contents. The complexity of MB content and texture change which are used to judge the importance of the MB. The unimportant MBs are divided into two slices based on edge match rule, which contributes to the error concealment in the decoder. The important slice is protected than the unimportant slice in the receiver so that the subjective quality of the reconstruction frame will be improved greatly. The proposed of adapter FMO scheme is to increase the error resilience of the encoded

video stream and contribute to the error concealment realization in the decoder. The adapter FMO strategy is suitable technique for the video transmission over low bandwidth.

Aziz et al. [3] present a technique to overcome the freezing frames problem on the mobile device and providing a smooth video playback over a wireless network. The frames in the streaming video will be splitted into four sub-frames on the server side and transmitted over Multiple-Input Multiple-Output (MIMO) by using the Multiple Descriptions Coding (MDC) technique. Where an initial delay time had been set between different channels to avoid the interruption on the sub-frames that are belong to the same frame. In case of the sub-frames that belonging to any subsequence are lost during the transmission, a reconstruction mechanism will be applied in the mobile device to recreate the missing pixels that are belongs to the missing sub-frames based on the average of the neighbouring pixels.

To overcome the transmission of each frame over MIMO and to increases the ability to handle long losses during the transmission over unreliable network. A splitting technique is proposed to deal with the sub-frames as equally important, by splitting the frames into sub-frames and cross them with another sub-frame from different sequence position.

The initial idea is been proposed in [4], where the frames been splitted into two sub-frames, where one sub-frame contains the even pixels and another contains the odd pixels. The combination of the sub-frame with another sub-frame from different sequences positions within the same transmission rate. The combined sub-frames will be streamed over a single wireless channel. In case of the frame being lost the available sub-frame on the mobile device will be reconstruct based on the surrounding pixels, while the maximum frames sequence lost that can be tolerated is half second.

The work has been extended to tolerate a maximum frame sequence lost up to six seconds (in worst cases), while the adaption mechanism allow us to stream up to one fourth (skipping three sub-frames) of the video frames to the mobile device according to the proposed technique. The reconstruction to the sub-frames in the video sequence will be measured by the Structural Similarity (SSIM) index.

4.3 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 the restricted processing power of the mobile terminals. The commonly used resolution is Quarter Common Intermediate Format (QCIF, 176 x 144 pixels) for mobile phones [8].

Mobile real time applications like video streaming suffer from high loss rates over the wireless networks [12] and the effect of that the mobile users may notice some sudden stop during the playing video, the picture is momentarily frozen. The frozen pictures could occur if a sequence of video frames is lost.

Distribute the frame's pixels as sub-frames over the streaming video is considered in this work by splitting each frame into four sub-frames [3], where each sub-frame contains one fourth of the main frame pixels, as shown in Figure 1. The crossing technique will be applied after splitting the frames as sub-frames and it will be crossed with other sub-frames that are from different frame sequence position.

During the interactive mode where the mobile clients request the connection to the video server, the server will start streaming the frames based on the frames splitting and frames crossing technique, as shown in Figure 2, 3 and 4, respectively.

Each video frame is splitted into s sub-frames, where s = 0,…S-1, where s is four sub-frames (A, B, C, D), as shown in Figures 1 and 3(a), respectively.

Each sub-frame contains different pixels information which makes it possible to implement the frames crossing technique among the frames groups to create the new frames crossing (FC).

The sequence of the video frames will be grouped on the streaming server according to the transmission rate per second as a frames group (FG), as shown in Figure 3(a), where g is the index of the frames group, g = 0,…, G-1.

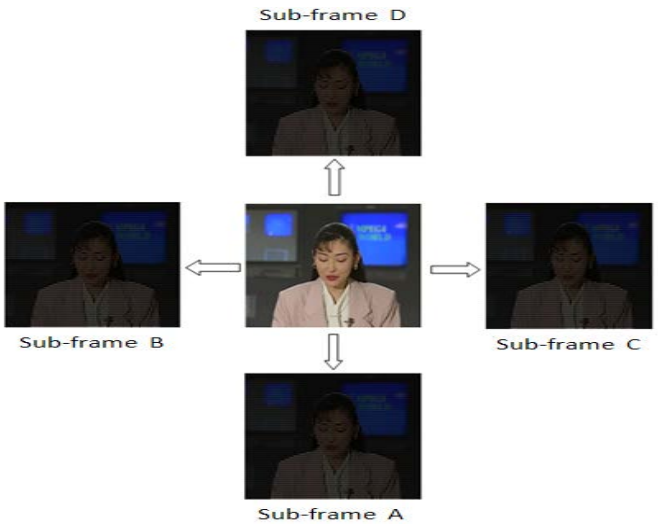


Figure 1: Snapshot of Akiyo frame splitting.

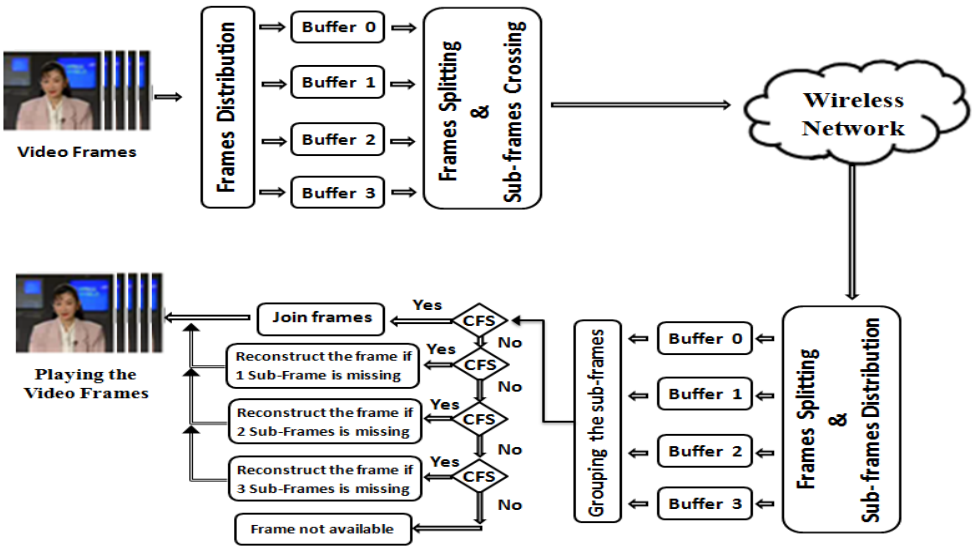


Figure 2: Streaming video as sub-frame crossing over wireless network.

To implement the frames crossing technique between different frames in different group where *i* is the index of the frames group crossing (FGC) where *i* = 0,…, S-1, where the sub-frames *s* of group *g* of the FGC *i* is obtained as

FGCi (g,s) = sF(G. ((s + i) mod S) + g , s), (1)

and are illustrated in Figures 3 and 4, respectively.

Crossing the sub-frames among the frames groups is required *s* buffers to queue the FGs, where the buffer size is equal to the frames rate, as shown in Figure 2. As an example, the first frames group FG0 will be queued in buffer 0, and the second FG1 will be queued in buffer 1, the third FG2 will queued in buffer 2, and the fourth FG3 will be queued in buffer 3. During the process of each buffer the next arrival group of frames, which is the fifth FG4, will be queued in buffer 0 and so on.

The transmission rate are considered in this work is 30 frames per second, where the frames group (FG) size will be 30 frames, during the arrival of the streaming video; the first 30 frames (FG0) will be splitted into four sub-frames, as shown in Figure 1 and 3(a). The same technique will be applied to the arrival of the second 30 frames (FG1) and so on.

When the first frame from the fourth group (FG3) of 30 frames arrived, the frames will be splitted into four sub-frames and the crossing technique will be applied immediately to distribute the frames pixels among the four groups in the streaming video, as shown in Figure 3.

		A	B	C	D
	0	$sF_{0,0}$	$sF_{0,1}$	$sF_{0,2}$	$sF_{0,3}$
	1	$sF_{1,0}$	$sF_{1,1}$	$sF_{1,2}$	$sF_{1,3}$
FG0	:	:	:	:	:
	:	:	:	:	:
	g	$sF_{g,0}$	$sF_{g,1}$	$sF_{g,2}$	$sF_{g,3}$
	g+1	$sF_{g+1,0}$	$sF_{g+1,1}$	$sF_{g+1,2}$	$sF_{g+1,3}$
	g+2	$sF_{g+2,0}$	$sF_{g+2,1}$	$sF_{g+2,2}$	$sF_{g+2,3}$
FG1	:	:	:	:	:
	:	:	:	:	:
	2g	$sF_{2g,0}$	$sF_{2g,1}$	$sF_{2g,2}$	$sF_{2g,3}$
	2g+1	$sF_{2g+1,0}$	$sF_{2g+1,1}$	$sF_{2g+1,2}$	$sF_{2g+1,3}$
	2g+2	$sF_{2g+2,0}$	$sF_{2g+2,1}$	$sF_{2g+2,2}$	$sF_{2g+2,3}$
FG2	:	:	:	:	:
	:	:	:	:	:
	3g	$sF_{3g,0}$	$sF_{3g,1}$	$sF_{3g,2}$	$sF_{3g,3}$
	3g+1	$sF_{3g+1,0}$	$sF_{3g+1,1}$	$sF_{3g+1,2}$	$sF_{3g+1,3}$
	3g+2	$sF_{3g+2,0}$	$sF_{3g+2,1}$	$sF_{3g+2,2}$	$sF_{3g+2,3}$
FG3	:	:	:	:	:
	:	:	:	:	:
	4g	$sF_{4g,0}$	$sF_{4g,1}$	$sF_{4g,2}$	$sF_{4g,3}$
	4g+1	$sF_{4g+1,0}$	$sF_{4g+1,1}$	$sF_{4g+1,2}$	$sF_{4g+1,3}$
	4g+2	$sF_{4g+2,0}$	$sF_{4g+2,1}$	$sF_{4g+2,2}$	$sF_{4g+2,3}$
FG4	:	:	:	:	:
	:	:	:	:	:
	5g	$sF_{5g,0}$	$sF_{5g,1}$	$sF_{5g,2}$	$sF_{5g,3}$
	5g+1	$sF_{5g+1,0}$	$sF_{5g+1,1}$	$sF_{5g+1,2}$	$sF_{5g+1,3}$
	5g+2	$sF_{5g+2,0}$	$sF_{5g+2,1}$	$sF_{5g+2,2}$	$sF_{5g+2,3}$
FGg-1	:	:	:	:	:
	:	:	:	:	:
	g-1	$sF_{g-1,0}$	$sF_{g-1,1}$	$sF_{g-1,2}$	$sF_{g-1,3}$

a. The sub-frames that are related to the original frame sequence

		A	B	C	D
	FC0	$sF_{0,0}$	$sF_{g+1,1}$	$sF_{2g+1,2}$	$sF_{3g+1,3}$
	FC1	$sF_{1,0}$	$sF_{g+2,1}$	$sF_{2g+2,2}$	$sF_{3g+2,3}$
FGC0	:	:	:	:	:
	:	:	:	:	:
	FCG-1	$sF_{G-1,0}$	$sF_{2G-1,1}$	$sF_{3G-1,2}$	$sF_{4G-1,3}$

b. The crossing frames position for FGC1

FGC1

	A	B	C	D
FCg+0	$sF_{g+1,0}$	$sF_{2g+1,1}$	$sF_{3g+1,2}$	$sF_{0,3}$
FCg+1	$sF_{g+2,0}$	$sF_{2g+2,1}$	$sF_{3g+2,2}$	$sF_{1,3}$
:	:	:	:	:
:	:	:	:	:
FC2G-1	$sF_{2G-1,0}$	$sF_{3G-1,1}$	$sF_{4G-1,2}$	$sF_{G-1,3}$

c. The crossing frames position for FGC2

FGC2

	A	B	C	D
FC2g+0	$sF_{2g+1,0}$	$sF_{3g+1,1}$	$sF_{0,2}$	$sF_{g+1,3}$
FC2g+1	$sF_{2g+2,0}$	$sF_{3g+2,1}$	$sF_{1,2}$	$sF_{g+2,3}$
:	:	:	:	:
:	:	:	:	:
FC3G-1	$sF_{3G-1,0}$	$sF_{4G-1,1}$	$sF_{G-1,2}$	$sF_{2G-1,3}$

d. The crossing frames position for FGC3

FGC3

	A	B	C	D
FC3g+0	$sF_{3g+1,0}$	$sF_{0,1}$	$sF_{g+1,2}$	$sF_{2g+1,3}$
FC3g+1	$sF_{3g+2,0}$	$sF_{1,1}$	$sF_{g+2,2}$	$sF_{2g+2,3}$
:	:	:	:	:
:	:	:	:	:
FC4G-1	$sF_{4G-1,0}$	$sF_{G-1,1}$	$sF_{2G-1,2}$	$sF_{3G-1,3}$

e. The crossing frames position for FGC4

Figure 3: The position of the sub-frames in the video sequence.

Figure 4: Snapshot for the Sub-frame crossing.

The crossing technique is implemented based on the frames crossing; where the frames crossing (FC) contains four different sub-frames from different FGs that belong to the same group. As an example, the first frame crossing FC0 will contains the sub-frame A from frame number 0, sub-frame B from frame number 30, sub-frame C from frame number 60, and sub-frame D from frame number 90, while the second FC1 will contains the sub-frame A from frame number 1, sub-frame B from the frame number 31, sub-frame C from frame number 61, and sub-frame D from frame number 91. In another way, the streaming video will be based on the sub-frames crossing and it will be transmitted as;

FC0(A0,B30,C60,D90),FC1(A1,B31,C61,D91),…,FC30(A30,B60,C90,D120),FC31(A31,B61,C91,D1),…,FC60(A60,B90,C120,D30),FC61(A61,B91,C1,D31),...,FC90(A90,B120,C30,D60),FC91(A91,B1,C31,D61),…,FC120(A120,B30,C60,D90),.., and so on, as shown in Figures 3 and 4 respectively.

The cost for implementing the proposed technique will be 3 seconds as an initial delay time, where the delay time is the time to queue FG0, FG1, FG2, for splitting and waiting for the fourth FG3, the time of the first frame from FG3 arrived it will be split and combine them with another frames from FG0, FG1, FG2 based on the proposed technique been described early. In this case we manage to distribute the frames pixels from different frame numbers and from different frames positions in the streaming video.

The crossing technique will be applied to all the frames in the video streaming sequence and it will be transmitted over a single channel. The reason behind that, if there is lost or dropped of sequence of frames from the streaming video and under different networks condition. The effect will be on at least one fourth of the sub-frames from the four different sub-frames that are in different positions. The quality of the video will be affected and it will be distributed on the streaming video frames.

After each frame has been received by the mobile device, a splitting frame technique will be applied. The sub-frames will be held in different buffers and according to the order they been splitted at the

server side, as shown in Figure 2. The sub-frames will be distributed to the relevant buffers and the combination of the sub-frames that are related to the same frame and according to their sequence positions based on switching between buffers to create the original frames sequences for the streaming video.

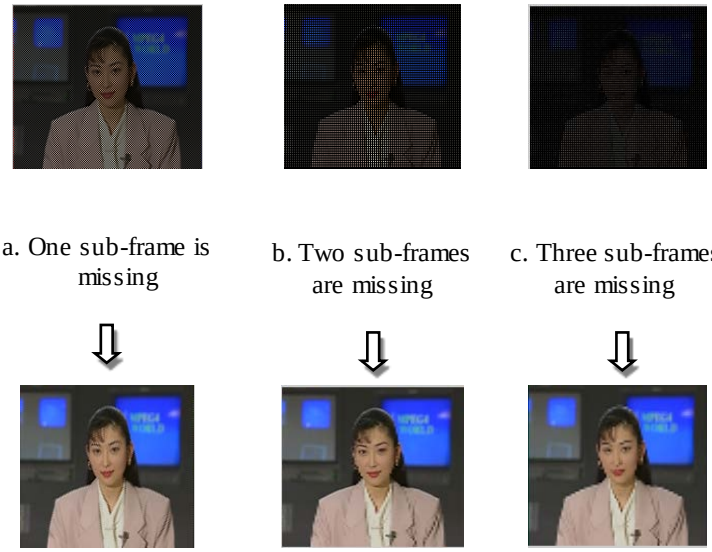


Figure 5: Akiyo snapshots of the missing and the reconstruction to the sub-frames.

The Check Frame Sequence (CFS) procedures will take place on the mobile device, to check the availability of the sub-frames. The CFS and the reconstruction mechanism are used to identify the missing sub-frames and to reconstruct the missing pixels from the frames by considering the following checking procedures [3], and as shown in Figure 5;

- 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 it is original shape.
- The second CFS will check if at least three sub-frames are available. If one sub-frame is missing, then the average of the neighbouring pixels will be calculated to replace the missing frame pixels.
- The third CFS will check if at least two sub-frames are available. If two sub-frames are missing, then the average of the neighbouring pixels will be calculated to replace the missing sub-frame pixels.
- The fourth CFS will check if at least one sub-frame is available. If three sub-frames are missing, then the average of the neighbouring pixels will be calculated twice, the first time to find the half of the frame and the second time to return the full frame to it is normal shape.

4.4 Rate Adaptation

Rate adaptation for streaming video is regarded as it is necessary mechanism to handle the network conditions, and the fluctuations of the network bandwidth.

The adaption rate for the sub-frames crossing technique should be considered carefully to avoid skipping the sub-frames that belonging to the same frame and with the consideration of the available bandwidth and network interruption to the streaming video. The adaption rate can be implemented by not considering the combination of the four sub-frames and transmitting either three or two or one sub-frame to the

mobile device and according to the following adjustments cases:

- 25% adjustment, the streaming server will skip only one sub-frame from the video frames sequence, as shown in Figure 5 (a).
- 50% adjustment, the streaming server will skip two sub-frames from the video frames sequence, as shown in Figure 5 (b).
- 75% adjustment, the streaming server will skip three sub-frames from the video frames sequence, as shown in Figure 5 (c).

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 75% from the frames pixels to avoid discard to the sub-frames that are related to the same video frame. The receiving sub-frames will be reconstruct to it is original frames, as shown in Figure 5.

4.5 Results and Discussion

In the normal situation, 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 mount of frames rate are arrived to start playing the video. While real time video streaming suffers from high loss rates over wireless networks [17], the result of 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.

The proposed technique is based on sub-frames crossing for the video test sequences Akiyo, Foreman, News, and Waterfall, as it is a well-known professional test sequences [19], with a transmission rate of 30 frames per second. The quality to the reconstructed sub-frames is expressed in terms of the Structural Similarity (SSIM) Index [18]. The SSIM index will measure the reconstructed video frames to the reference frames, as shown in Figures 6, 7, and 8 respectively.

Considering the same losing frame sequence in [3], where the

number of frames are lost are 20 frames as light lost rate from the streaming video, then the effect of losing frames will be distributed on the streaming sequence and the affect will be on 80 frames, as these frames will loss one sub-frame. As an example, if the frame losses is started from frame 121 to 140, then the effect of losing one sub-frame will affect the frames sequence from 121 - 140, 151 - 170, 181 - 200, and from 211 - 230, as the losses of these frames are fall in the same crossing group. The frames that lost the sub-frame it will be reconstructed and therefore, the quality level of the frames will be affected.

If the numbers of frames are lost are 40 frames as medium lost rate from the streaming video, then the effect of losing frames will be distributed on the streaming sequence and the affect will be on 120 frames, as some frames will lose one sub-frame while others will lose two sub-frames. As an example, if the loss of frames starts from frame 121 to 160 then the effect of losing one sub-frame from 131-150, 161-180, 191-210, 221-240. While the following frames sequence will lose two sub-frames will affect the frames 121-130, 151-160, 181- 190, 211-220. Therefore, the quality level of the video will be distributed on the video sequence after been reconstructed as some video frames loss one sub-frame and others will loss two sub-frames.

If the numbers of frames are lost are 60 frames as high lost rate from the streaming video, then the effect of losing frames will be distributed on the streaming sequence and the effect will be on 120 frames. As an example, if the falls of frame losses are started from frame 121 to 180, then the effect of losing sub-frames will affect the frames from 121 to 240 as all the effected frames will loss two sub-frames. The receiving sub-frames will be reconstructed on the mobile devices to return the missing pixels for each frame and played on the mobile screen with less quality than the original frames. The losses duration can be handle in this technique is up to six seconds, as shown in Figure 3. If the losses occur in the FGC1, FGC2, FGC3, FGC4, FGC5, and FGC6, the mobile device will received the following sequence of one sub-frame from 0 until 239, as these sub-frames are received by FGC0 and FGC7.

Original frame

The reconstruction for one sub-frame missing

SSIM : 0.955

Original frame

The reconstruction for one sub-frame missing

SSIM : 0.948

The reconstruction for two sub-frames missing

SSIM : 0.929

The reconstruction for three sub-frames missing

SSIM : 0.911

The reconstruction for two sub-frames missing

SSIM : 0.941

The reconstruction for three sub-frames missing

SSIM : 0.907

a. Akiyo

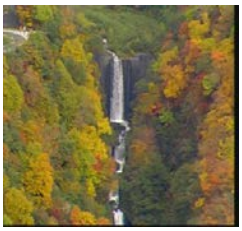
b. Foreman

Original frame

The reconstruction for one sub-frame missing

SSIM : 0.939

Original frame



The reconstruction for one sub-frame missing

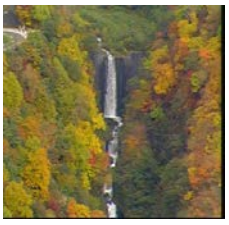
SSIM : 0.982

The reconstruction for two sub-frames missing

SSIM : 0.923

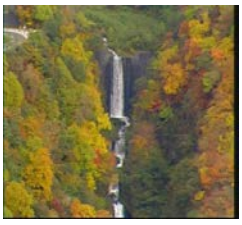
The reconstruction for three sub-frames missing

SSIM : 0.874



The reconstruction for two sub-frames missing

SSIM : 0.972



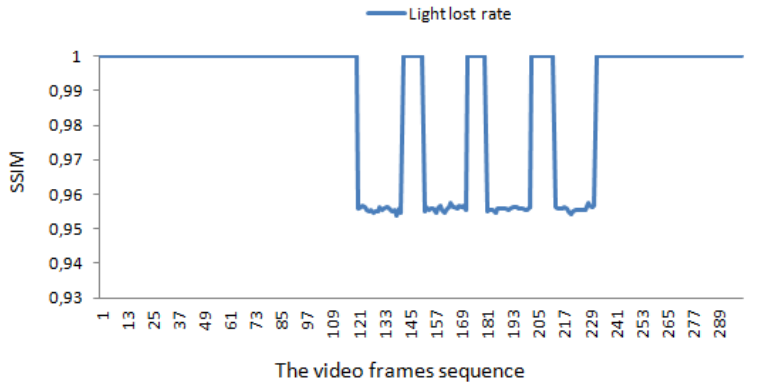
The reconstruction for three sub-frames missing

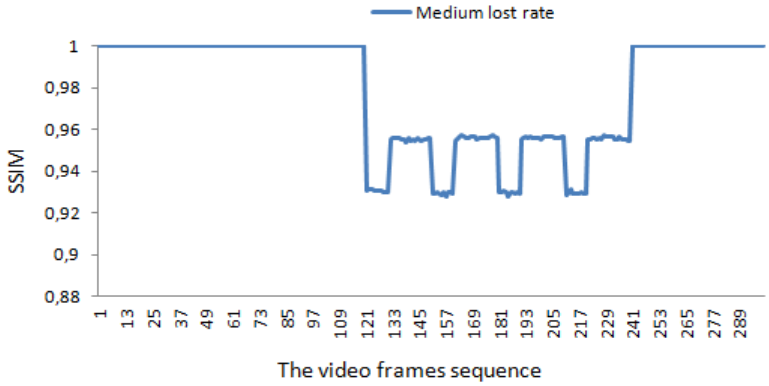
SSIM : 0.945

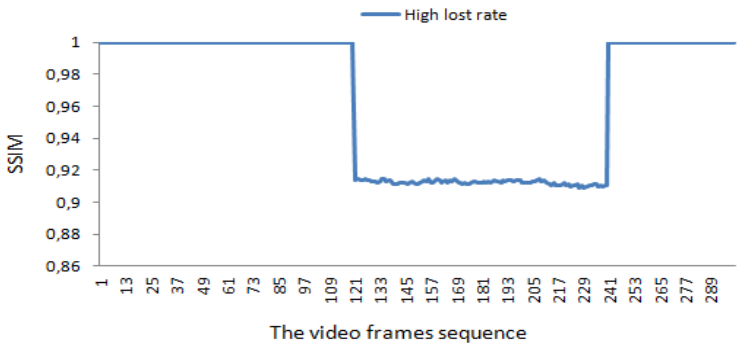
c. News

d. Waterfall

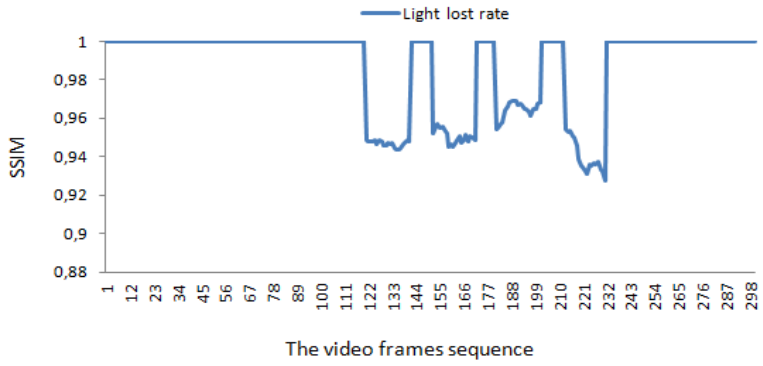
Figure 6: The SSIM for the frame number 140.

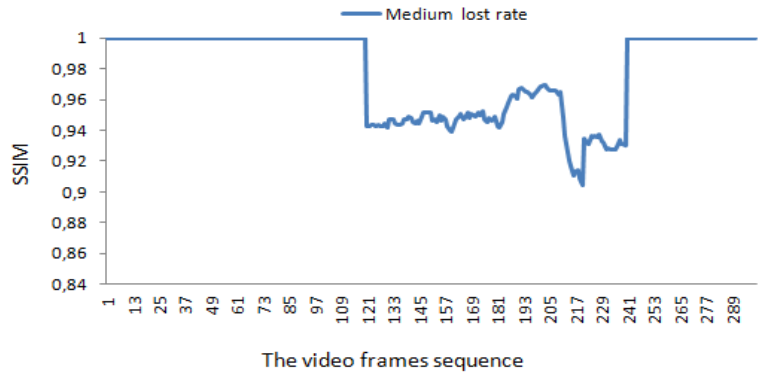

Light lost rate
SSIM
1
0,99
0,98
0,97
0,96
0,95
0,94
0,93
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence

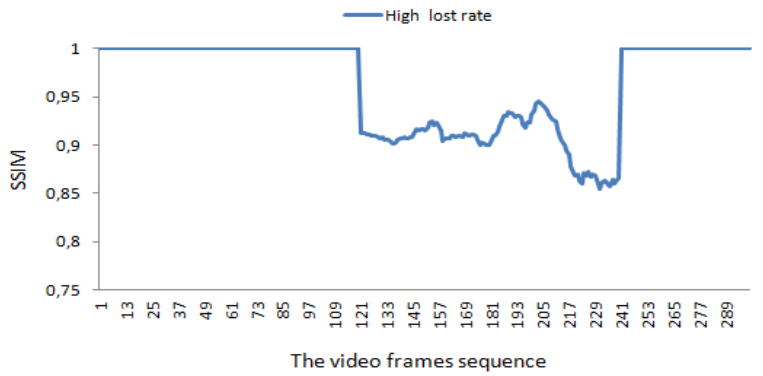

Medium lost rate
SSIM
1
0,98
0,96
0,94
0,92
0,9
0,88
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence


High lost rate
SSIM
1
0,98
0,96
0,94
0,92
0,9
0,88
0,86
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence

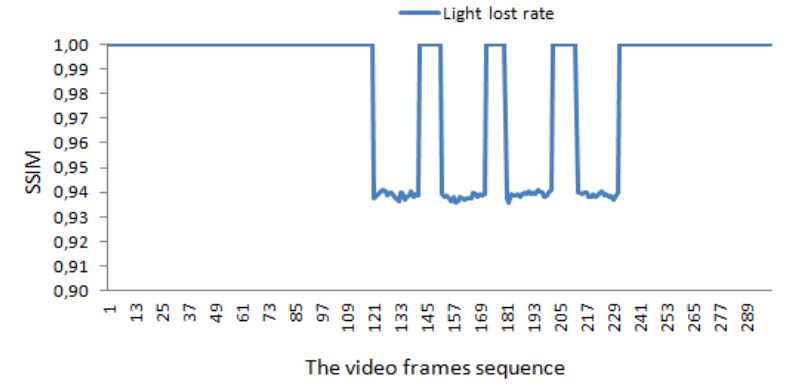
a. Akiyo

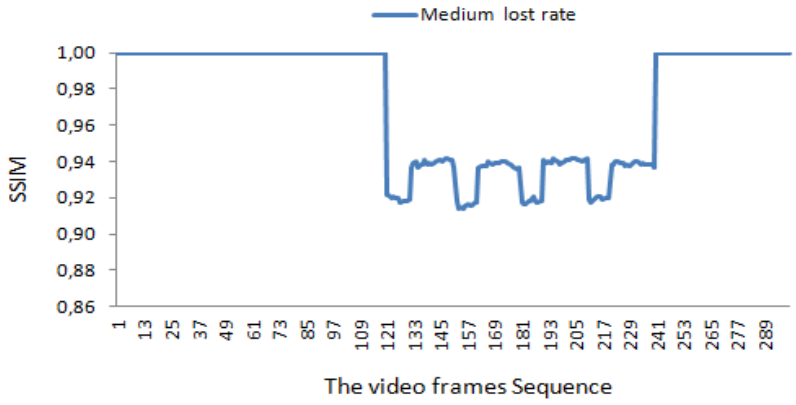

Light lost rate
SSIM
1
0,98
0,96
0,94
0,92
0,9
0,88
1 12 23 34 45 56 67 78 89 100 111 122 133 144 155 166 177 188 199 210 221 232 243 254 265 276 287 298
The video frames sequence

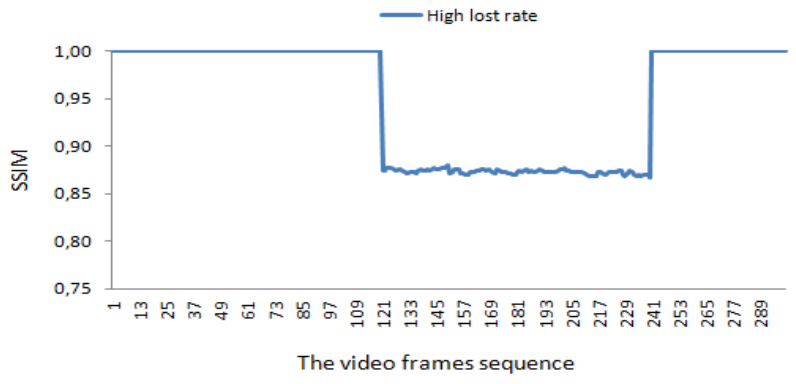

Medium lost rate
SSIM
1
0,98
0,96
0,94
0,92
0,9
0,88
0,86
0,84
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence


High lost rate
SSIM
1
0,95
0,9
0,85
0,8
0,75
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence

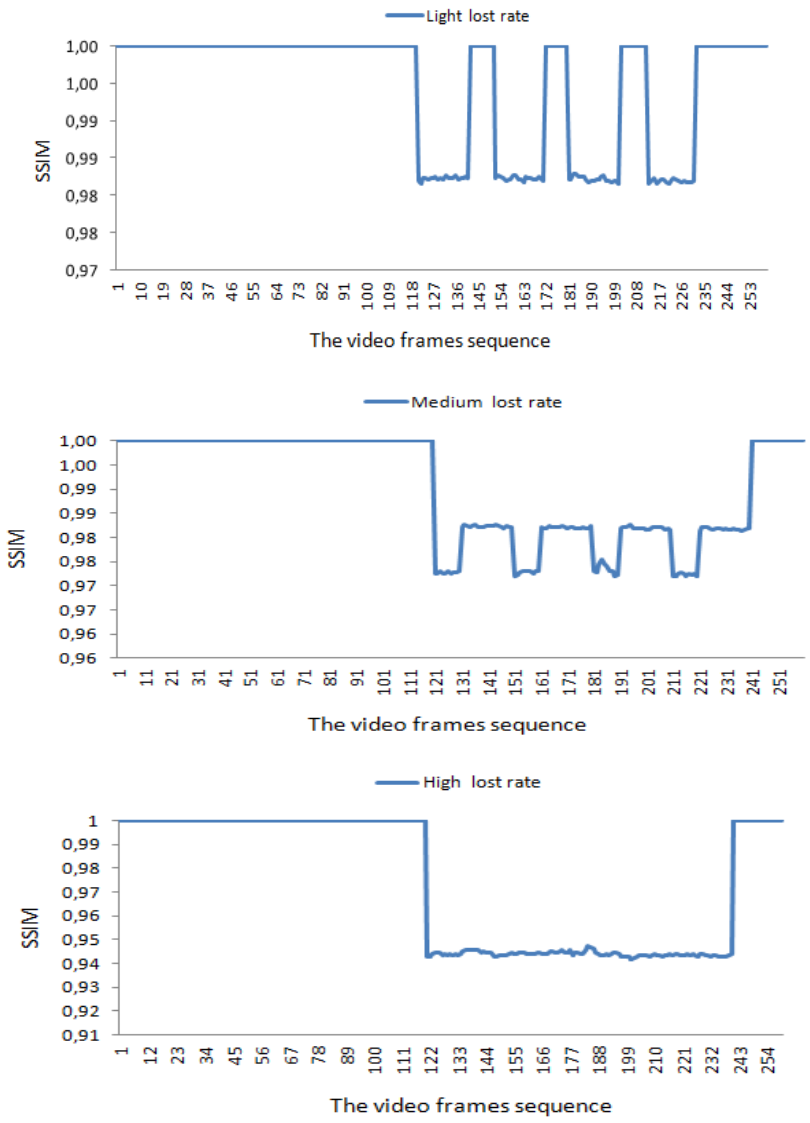
b. Foreman


Light lost rate
SSIM
1,00
0,99
0,98
0,97
0,96
0,95
0,94
0,93
0,92
0,91
0,90
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence


Medium lost rate
SSIM
1,00
0,98
0,96
0,94
0,92
0,90
0,88
0,86
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames Sequence

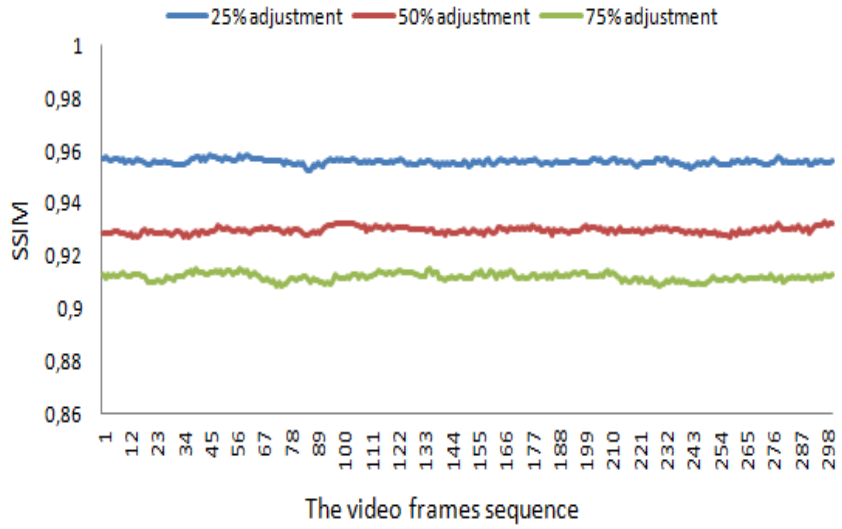

High lost rate
SSIM
1,00
0,95
0,90
0,85
0,80
0,75
1 13 25 37 49 61 73 85 97 109 121 133 145 157 169 181 193 205 217 229 241 253 265 277 289
The video frames sequence

c. News

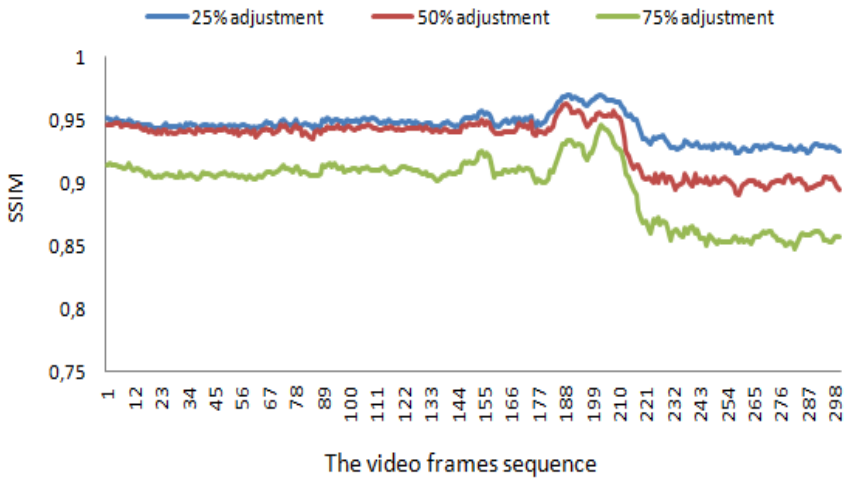


d. Waterfall

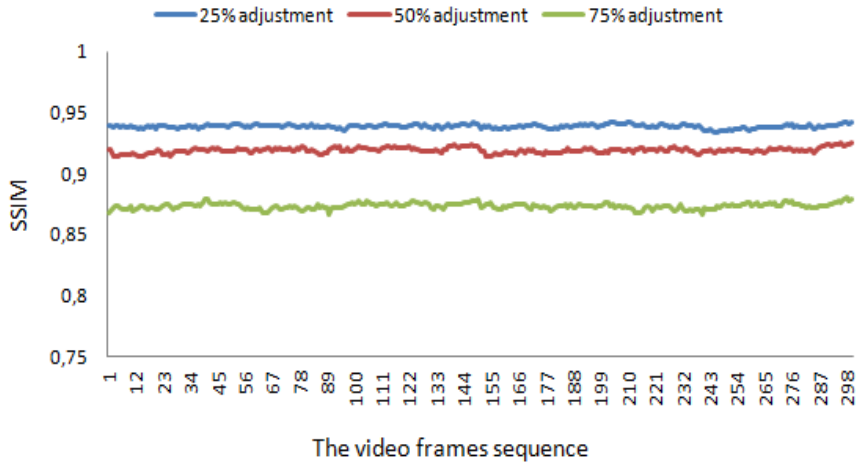
Figure 7: The SSIM for video frames after the lost been distributed and reconstructed to the sub-frames.



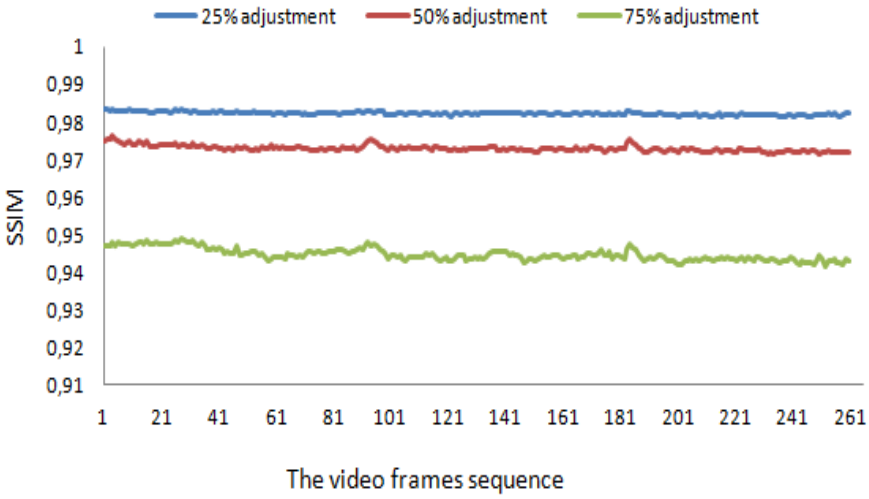
a. Akiyo



b. Foreman



c. News



d. Waterfall

Figure 8: The SSIM for the reconstruction sub-frames for the adaption rate to the video frames sequence.

The adaption rate is also considered in this paper, where the server can skip either one, or two, or three sub-frames, where the quality level of the video will be affected according to the adaption rate, as shown in Figure 8.

Skipping three sub-frames shows low quality than skipping two or one sub-frame. The Waterfall video shows better results as the pixels of the video frames have similar data where the reconstruction mechanism did not been effected that much, while the News video is been effected highly by the reconstructions mechanism as it is quite motion video and it can be seen clearly in Figure 6.

4.6 Conclusion

In this paper, we proposed a sub-frames crossing technique to distribute the pixels as sub-frames in different positions in the sequence of the streaming video by combining it with other sub-frames from different positions. The idea behind that is to eliminate the losses of the complete single frame and allow at least one fourth of the frame (one sub-frame) to be received by the mobile device. The receiving sub-frames will be reconstructed based on the neighboring pixels to replace the missing pixels.

From the results, it is shown that our proposed technique provides a promising direction for eliminating the frozen picture on the mobile screen, that been caused by missing frames from the streaming sequence. Adjusting the number of frames according to the bandwidth changes is highly needed to reduce the amount of data to be transmitted to the mobile device in a congested network.

However, the quality of the played video is degraded and it depends on the number of frames that are lost or skipped. The numbers of buffers are needed will be equivalent to the number of crossing group, while the initial delay time it needed to implement the crossing technique.

References

[1] G. Bai, and C. Williamson, "The Effects of Mobility on Wireless Media Streaming Performance," Proc. of the Wireless Networks and Emerging Technologies (WNET 04), July 2004, pp. 596-601.

[2] 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," Proc. of the 6th IEEE International Conference on Computer and Information Technology (CIT 06), Sept. 2006, pp. 177-177,doi: 10.1109/CIT.2006.158.

[3] H. M. Aziz, M. Fiedler, H. Grahn, and L. Lundberg, "Streaming Video as Space – Divided Sub-Frames over Wireless Networks," Proc. of the 3rd Joint IFIP Wireless and Mobile Networking Conference (WMNC 10), Oct. 2010, pp. 1-6, doi: 10.1109/WMNC.2010.5678760.

[4] H. M. Aziz, H. Grahn, and L. Lundberg, "Sub-Frame Crossing for Streaming Video over Wireless Network," Proc. of the 7th International Conference on Wireless On-demand Network Systems and Services (WONS 10), Feb. 2010, pp. 53 - 56, doi: 10.1109/WONS.2010.5437132.

[5] H. Zhu, H. Wang, I. Chlamtac, and B. Chen, "Bandwidth Scalable Source-Channel Coding for Streaming Video over Wireless Access Networks," Proc. of Wireless Networking Symposium (WNCG 03), Oct. 2003.

[6] H. Luo, M.-L., Shyu, and S.-C. Chen, "An End-to-End Video Transmission Framework with Efficient Bandwidth Utilization," Proc. of the IEEE International Conference on Multimedia and Expo (ICME 04), June 2004, pp. 623-626, doi: 10.1109/ICME.2004.1394269.

[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, Dec. 2003, pp. 1566-1574, doi: 10.1109/JSAC.2003.814863.

[8] M. Ries, O. Nemethova, and M. Rupp, "Performance Evaluation of Mobile Video Quality Estimators," Proc. of the European Signal Processing Conference (EUSIPCO 07), Sept. 2007, pp. 159-163.

[9] P. Antoniou, V. Vassiliou, and A. Pitsillides, "ADIVIS: A Novel Adaptive Algorithm for Video Streaming over the Internet," Proc. of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07), Dec. 2007, doi: 10.1109/PIMRC.2007.4394583.

[10] R. Weber, M. Guerra, S. Sawhney, L. Golovanvsky, and M. Kang, "Measurement and Analysis of Video Streaming Performance in Live UMTS Networks," Proc. of the 13th International Symposium on Wireless Personal Multimedia Communications (WPMC 06), Sept. 2006, pp. 1-5.

[11] S. Cen, C. Pu, and R. Staehli, "A Distributed Real-time MPEG Video Audio Player", Proc. of the 5th International Workshop on Network and Operating System Support of Digital Audio and Video, LNCS, 1995, pp. 142-153, doi: 10.1007/BFb0019263.

[12] T. Nguyen, P. Mehra, and A. Zakhor, "Path Diversity and Bandwidth Allocation for Multimedia Streaming," Proc. of the International Conference on Multimedia and Expo (ICME 03), July 2003, pp. 1-4.

[13] T.Y. Huang, "Region of Interest Extraction and Adaptation in Scalable Video Coding," Proc. of the 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 10), Aug. 2010, pp. 2320-2323, doi: 10.1109/FSKD.2010.5569822.

[14] X. Cao, G. Bai, and C. Williamson, "Media Streaming Performance in a Portable Wireless Classroom Network," Proc. of IASTED European Workshop on Internet Multimedia Systems and Applications (EuroIMSA´05), Feb. 2005, pp. 246-252.

[15] X. Zhu, and B. Girod, "Video Streaming over Wireless Networks," Proc. of the European Signal Processing Conference (EUSIPCO 07), Sept. 2007, pp. 1462-1466.

[16] X. Wang, and X. Tu, "Adaptive FMO Strategy for Video Transcoding," Proc. of the International Conference on Communications, Circuits and Systems (ICCCAS 09), July 2009, pp. 540 - 544, doi: 10.1109/ICCCAS.2009.5250462.

[17] Y. Wang, A. R. Reibman, and S. Lin, "Multiple Description Coding for Video Delivery," Proc. of the IEEE Journal, vol. 93, Dec. 2004 pp. 57-70, doi: 10.1109/JPROC.2004.839618.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Transactions on Image Processing, vol. 13, April 2004, pp. 600-612, doi: 10.1109/TIP.2003.819861.

[19] http://trace.eas.asu.edu/yuv/index.html (visited, 1/11/2011)

[20] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” Proc. of the IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, Sept. 2007, pp.1103-1120, doi: 10.1109/TCSVT.2007.905532.

CHAPTER FIVE

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 delays, which could have negative effects on the video streaming and the viewer will perceive a frozen picture for certain durations due to loss of frames. In this study, 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 objective of streaming the video in this way is to avoid the losses of a sequence of neighbouring frames in case of a long sequence interruption. We evaluate our approach by using a user panel and Mean Opinion Score (MOS) measurements; where the users observe three levels of frame losses. The results show that our technique significantly improves the smoothness of the video on the mobile device in the presence of frame losses, while the transmitted data are only increased by almost 9 % (due to reduced time locality).

Keywords

Streaming Video, Frozen Pictures, Interleaving, Switching Frames, Mean Opinion Score

5.1 Introduction

With the rapid development of video coding and wireless communication technologies, video streaming has become very popular to mobile users [11]. The H.264/AVC coding standard has been introduced to achieve high compression of the video stream. 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 end viewer's perceived quality [22].

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 the Real Time Protocol (RTP) to detect packet losses [9]. 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 packet losses. As a result, these packet losses between the video sender and the receiver significantly degrade the quality of the received video [2].

Real time video streaming is particularly sensitive to frame losses and delays since the frames must arrive at the mobile device before their playout time with enough time to decode and display the frames [8]. The available bandwidth of a wireless channel varies with time, and that can potentially result in loss of frames, called outage [1, 23]. If the receiver buffer runs out of packets, the playout of the video is interrupted until the buffer has received as many frames that are required to decode and continue to display the video sequence on the mobile device [13]. Too long outages (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) [13, 16]. Even a low amount of frame losses can result in a severe degradation of quality as perceived by the users [12, 24]. 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 [19, 21].

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

In the 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 encode the video. Therefore, the size of the streaming video may increase due to interleaving. Consequently, interleaving will enhance 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 one after the 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 on to the display screen in the original order. This means that we need to introduce extra delay time due to interleaving.

In this paper, we quantify the gain in user-perceived quality caused by frame interleaving for video transmissions over wireless channels with potential frame losses. We also quantify the increased sized of the data encoding the video due to reduced time locality. Finally, we calculate the relation between the maximum outage that can be handled and the size of the receiver buffer. It turns out that, TIRS on average can handle network outages that are longer than the (jitter) buffer on the receiver side.

5.2 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 [6, 7, 18].

Cai and Chen [5] 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 packet losses on ATM networks.

Schierl et al. [23] presented a streaming system that utilizes interleaved transmission for real-time H.264/AVC video and audio in wireless environments. In their approach, they consider the audio as 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.

An 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. [17]. The protocol combines features of MPEG-4 data partitioning, FEC, and interleaving techniques. Adaptive FEC transmission based on efficient feedback for 1-to-*n* multicast communication scenario is also considered in their study.

Tsai et al. [25] proposed a technique to disperse 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 handle the impact of packet disordering. Path interleaving aims at striping two or more FEC blocks to multiple paths to spread the burst of packet losses to different blocks.

The above authors used feedback techniques and retransmission mechanisms for error recovery for lost packets. Retransmission due to 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. There will be overhead on the server side as well as in the wireless network, and the overhead will depend on the network condition and the amount of packets lost. Therefore, we propose the TIRS technique to avoid the use of feedback techniques and retransmission mechanisms.

5.3 The Interleaving Distance Algorithm

Claypool and Zhu [7] proposed video interleaving as a repair technique to ameliorate the effect of frame losses from the streaming video. They re-sequence the frames in a video stream at the sender side, and stream them through the wireless network to the receiver side. They proposed an Interleaving Distance Algorithm (IDA) for recording television video programs before they are encoded by using MPEG-1. The decoding tool used is the Berkeley MPEG-2 player. In their study, the interleaving distances 2 and 5 were chosen: 2 for short distance interleaving and 5 for long distance interleaving. The interleaving distance determines how long time consecutive frames are spread out in the video stream.

The Group of Pictures (GOP) length [7] was used as a basis for their interleaving technique. For a GOP of 9 frames (used in their paper) and an interleaving distance of 2, they first encode and transmit the first 9 even frames 2, 4, 6, 8, 10, 12, 14, 16, 18, in the original video, and then they encode and transmit the first 9 odd frames 1, 3, 5, 7, 9, 11, 13, 15, 17, as shown in Figure 1.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 1819 20 21 22 23 24 25 26 27 28 29 30 ...

a. Original sequence

1 3 5 7 9 11 13 15 17 2 4 6 8 10 12 14 16 18 ...

b. Interleaving distance 2

1 4 7 10 13 16 19 22 25 2 5 8 11 14 17 20 23 26 3 6 9 12 15 18 21 24 27 ...

c. Interleaving distance 3

1 5 9 13 17 21 25 29 33 2 6 10 14 18 22 26 30 34 3 7 11 15 19 23 27 31 35 4 8 12 16 20 24 28 32 36 ...

d. Interleaving distance 4

1 6 11 16 21 26 31 36 41 2 7 12 17 22 27 32 37 42 3 8 13 18 23 28 33 38 43 4 9 14 19 24 29 34 39 44 5 10 15 20 25 30 35 40 45 ...

e. Interleaving distance 5

Figure 1: Frame sequences with different interleaving distances and GOP length 9.

After that they 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 interleaving of IDA is shown in Figure 1 for different interleaving distances. In this case, we are interleaving 30 frames. Figure 1(e) corresponds to the case with interleaving distance 5 and GOP length of 9. The interleaving mechanism in IDA changes the frame positions in the video before it is compressed and streamed over the wireless network. The number of buffers on the sending and receiving side needed to hold the frames is equal to the interleaving distance, as shown in Figure 2. Frames in the video sequence will be split according to the interleaving distance N. If N= 3, as shown in Figure 1 (c), the frames will be forwarded to three different buffers. The first buffer will hold frames number 1, 4, 7, 10, 13, 16, 19, 22, 25, the second buffer will hold frames number 2, 5, 8, 11, 14, 17, 20, 23, 26, and the third buffer will hold frames number 3, 6, 9, 12, 15, 18, 21, 24, 27. The size of each buffer is equivalent to the length of the GOP.

The video frames will be compressed and streamed based on switching between the buffers to create the frame sequence interleaving order. For interleaving distance N=3, it sends the frame group from the first buffer, followed by the frame group from the second buffer, followed by the frame group from the third buffer, followed by the frame group from the first buffer and so on. For instance, we transmit the first 27 frames in the following order 1, 4, 7, 10, 13, 16, 19, 22, 25, 2, 5, 8, 11, 14, 17, 20, 23, 26, 3, 6, 9, 12, 15, 18, 21, 24, 27.

In the normal case, the mobile device will start playing the video as soon as the frames are received (with the possible exception of jitter buffers). When using interleaving, frames will arrive in a different order according to IDA. When the mobile device starts receiving the interleaved video stream, it will decode the video using the H.264 ffmpeg decoder [27]. The received frames will be forwarded to the appropriate buffers, i.e. the odd buffer or the even buffer when using interleaving distance 2. The general case with interleaving distance N is shown in Figure 2.

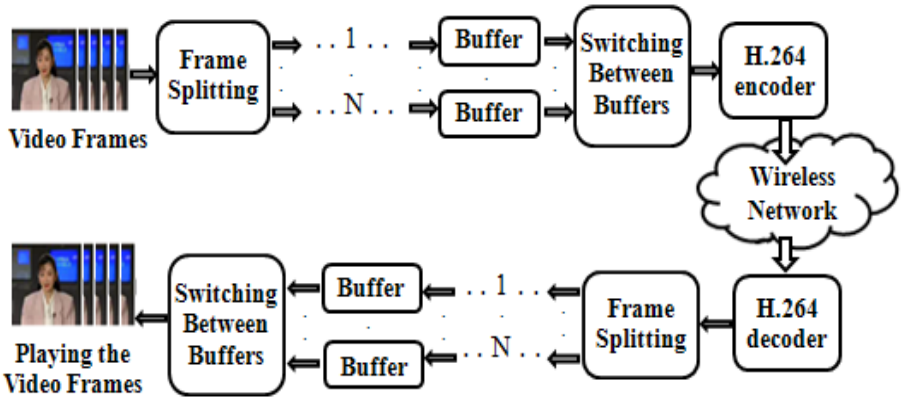


Figure 2: Streaming video as IDA over wireless network.

The latency time for playing the video is based on the interleaving distance and the GOP. The waiting time will be (GOP * (*N*-1)) / transmission rate**,** as an example, if we assume that the IDA interleaving distance is 3 as shown in Figure 1(c), and then the waiting time before start playing the video is 0.6 second given a video with 30 frames per second. The longer interleaving distances that are applied, the more is the waiting time is needed before we can start playing the video on the mobile device. In Section 5.6 we investigate the video size using IDA and H.264 for interleaving distances 2, 3, 4, and 5.

5.4 The Time Interleaving Robust Streaming Technique

In this section, we introduce the Time Interleaving Robust Streaming (TIRS) technique to stream video frames. The idea behind TIRS is to avoid frozen pictures on the mobile device when streaming over unreliable networks. This can be done by streaming the video as groups of even and odd frames. The TIRS technique will be applied before the video has been encoded and compressed by the H.264 encoder. TIRS can be implemented for different interleaving times (Δt) to distribute frames to different positions in the streaming sequence and to avoid a consecutive loss of frames in the streaming video.

In following example Δt = 1 second. In TIRS a sequence of even frames is follow by a sequence of odd frames. We assume 30 frames/second and since Δt = 1 second, we get a frame group (FG) of 30 frames (Δt multiplied with the transmission rate). For each FG we have two subgroups: Fe and Fo, which are the even and odd frames belonging to the same FG as shown in Figure 3.

The second step is to create the streaming video based on the interleaving technique. The even and odd frames will be grouped and distributed in different position in the streaming sequence; we start with Fe followed by Fo.

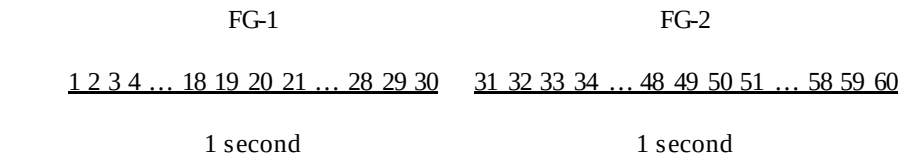
The even frames (Fe) from FG-1 and FG-2 will be grouped and streamed followed by a sequence of odd frames (Fo) from FG-1, FG-2, FG-3 and FG-4, and followed by a sequence of even frames (Fe) from FG-3, FG-4, FG-5 and FG-6 and so on. For example, the frames will be transmitted in the following order 2, 4, ... , 5 8, 60, 1, 3, … , 117, 119, 62, 64,…, 178, 180, 121, 123,…, as shown in Figure 4. The reason why we send four groups of odd frames instead of two groups is that we want to minimize the reduction of time redundancy. This will increase the compression rate when using H.264. There are two buffers: the even buffer contains the Fe frames and the odd buffer contains the Fo 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 and before it is been encoded by H.264 encoder. After that, the time interleaving streaming sequence will be streamed through the wireless network as shown in Figure 4 and 5. When the mobile device starts receiving the video, the H.264 decoder will decode the video and the frames will be split into even and odd frames and forwarded to the two buffers.

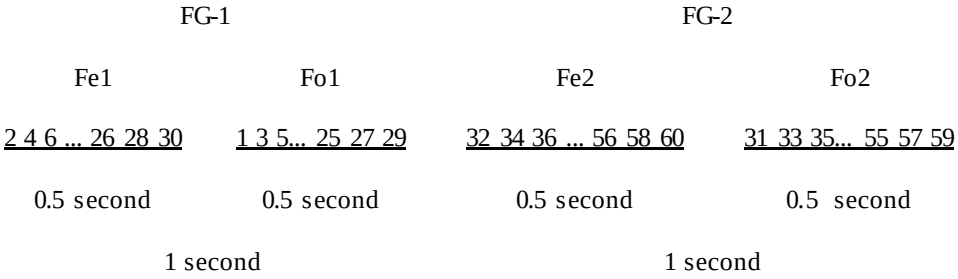
At start-up, the even frames will be delayed by Δt waiting for the odd frames to arrive to the odd buffer. The number of frames that has arrived to the even buffer is $V = \Delta t * R$, where R is the number of frames per second received by the buffer. This means that the size of each buffer is V.

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

A Check Frame Sequence (CFS) will be used to check the missing frames from the streaming video [3]. 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 [20].



a. Original stream



b. Frames interleaving within the same frame group (FG)

Figure 3: Video streaming sequence.

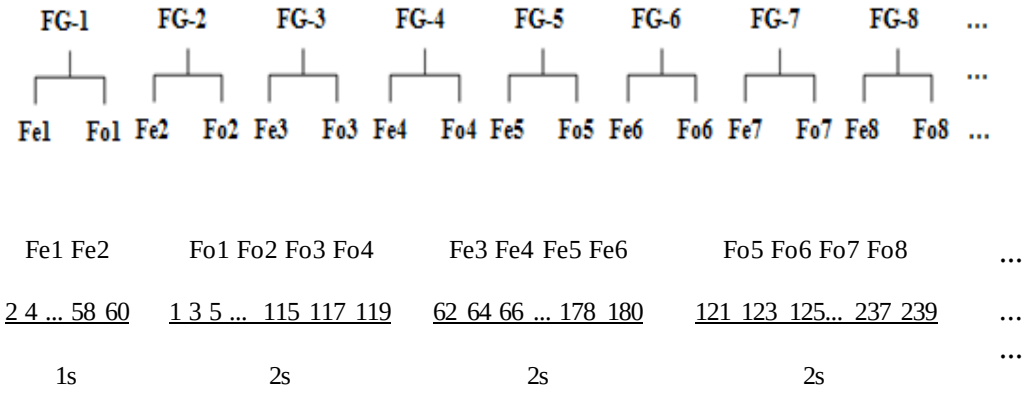


Figure 4: The proposed TIRS technique, where $\Delta t = 1$ second.

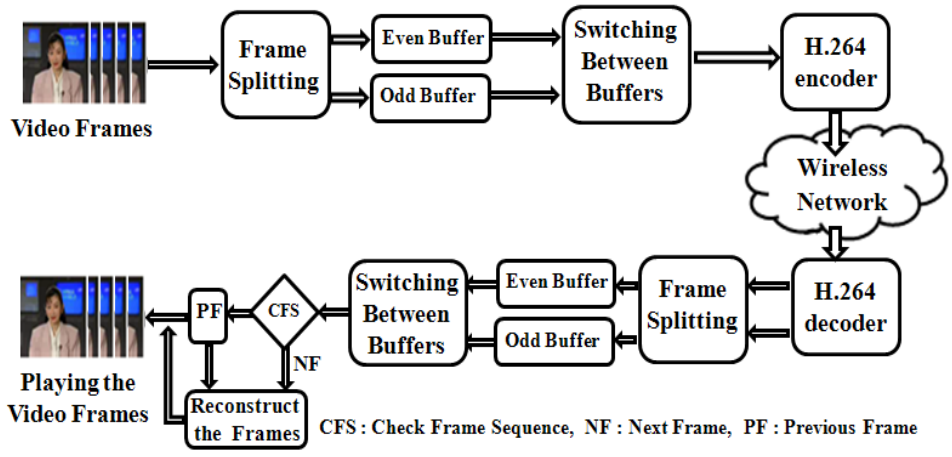
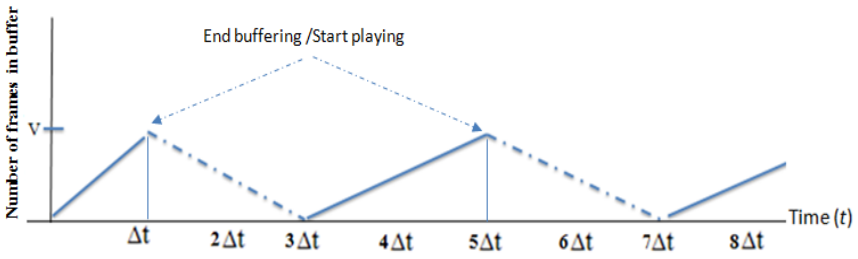
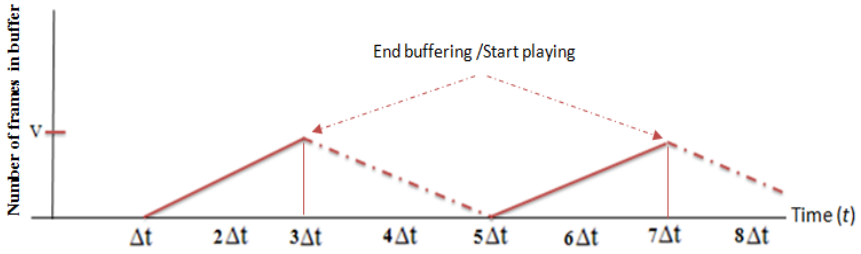


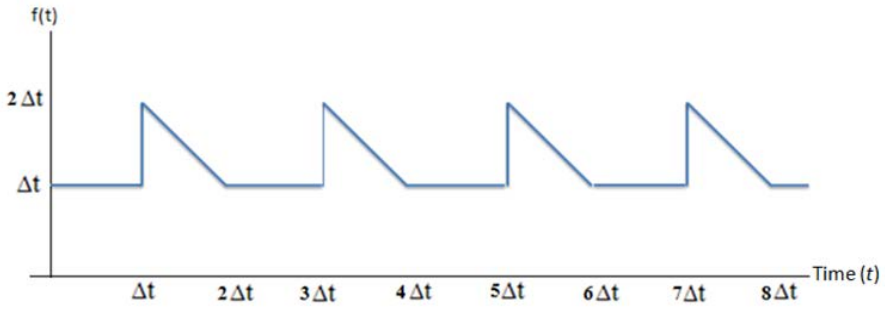
Figure 5: Streaming video using TIRS over wireless networks.



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.

5.5 The Effect of Losses on the Interleaving Frames

Streaming video over unreliable network could have a significant effect on the video quality, especially when a consecutive sequence of frames is lost. The playout of a video starts when there are available frames in the buffer. In the TIRS case, we need to compensate for the delay due to frame group (FG) interleaving. Interleaving has the advantage that we can use frame interpolation to tolerate loss of frame sequences up to a certain length. By using long interleaving intervals (e.g., Δt =2 seconds instead of Δt =1 second), we can obviously tolerate losses of longer frame sequences, which is good in case of long 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 sequences (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. This means that the video will be shown at the receiver 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 $3\Delta t$, then the lost odd frames can be recreated by interpolating the receiving even frames from 0 to Δt.

If the outage starts at $t = 0$ until $1.5\Delta t$, 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 stage of the streaming video and its greater than Δt, then we will face a frozen video, but if the outage occurs between Δt and $3\Delta t$ then there is possibility to interpolate the missing frames. If the outage occurs between $2\Delta t$ and $4\Delta t$ and it is greater than Δt, we will face a frozen picture.

5.6 The Effect of Interleaving on the File Streaming Size

Compressing the video frames based on IDA and TIRS by using the H.264 ffmpeg encoder [27] is affected by the amount of redundancy that can be removed. When the frame positions change, the temporal locality decreases which leads to increased size of the video stream for both techniques. IDA and TIRS have been applied on the video test sequences Akiyo, Foreman, News, Waterfall, and Football [26]. The chosen videos are well known as professional test sequences that have different characteristics, with a transmission rate of 30 frames per second, and each video has a 10 seconds duration with QCIF (176×144 pixels) resolutions.

5.6.1 The Effect of IDA on the File Streaming Size

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

The size increase varies from one video to another since each video has different features which affect the temporal locality, and thus also the size, differently. As the interleaving distance in IDA increases the temporal locality is in most cases reduced, and the size of the video therefore increases. The exception is the Foreman video, where the size for IDA-5 is somewhat smaller than the size using IDA-4; this is a coincidence.

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

Videos	**Akiyo**	**Foreman**	**News**	**Waterfall**	**Football**
Original	2886120	7151874	4115786	6679052	8524180
No. of packets	1961	4859	2796	4537	5791
IDA-2	**3093878**	**7348202**	**4465012**	**7243318**	**8597744**
No. of packets	2102	4992	3033	4921	5841
Increase	**7.19%**	**2.73%**	**8.48%**	**8.44%**	**0.86%**
IDA-3	**3227560**	**7429644**	**4637112**	**7474446**	**8622794**
No. of packets	2193	5047	3150	5078	5858
Increase	**11.83%**	**3.88%**	**12.66%**	**11.90%**	**1.16%**
IDA-4	**3295714**	**7461896**	**4726700**	**7617466**	**8634452**
No. of packets	2239	5069	3211	5175	5866
Increase	**14.19%**	**4.33%**	**14.84%**	**14.05%**	**1.29%**
IDA-5	**3385274**	**7450040**	**4778734**	**8711264**	**8667018**
No. of packets	2300	5062	3246	5918	5888
Increase	**17.29%**	**4.16%**	**16.10%**	**30.42%**	**1.67%**

It is worth noticing that the increase of size for the highly dynamic Football video is not bigger than the increase of size for the more static videos. It seems that even if the dynamic videos clearly contain less temporal locality, this does not mean that the relative reduction of temporal locality (and thus relative increase of size) due to interleaving is larger for dynamic videos than for more static videos.

5.6.2 The Effect of TIRS on the File Streaming Size

The effect of different time interleaving on the streaming file size is shown in Table 2, where the value of Δt considered are 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 shown in Figure 4. For Δt = 2 seconds, 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 is applied before the videos are compressed by the H.264 ffmpeg encoder [27]. The sizes of the compressed video clips increase differently and for different videos, as frames are moved to different scene position, affecting the temporal locality as shown in Table 2.

The changes in the streaming file size for the three interleaving times are rather similar to each other except the News video as the 2 seconds interleaving show much less increases in the streaming size. This is probably due to a coincidence, i.e. the frames happen to be moved to a position where there is similar data and the encoder will hence be able to remove more redundant data.

It is again worth noticing that the increase of size for the highly dynamic Football video is not bigger than the increase of size for the more static videos. It seems that even if the dynamic videos clearly contain less temporal locality, this does not mean that the relative reduction of temporal locality (and thus relative increase of size) due to interleaving is larger for dynamic videos than for more static videos.

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

Videos	**Akiyo**	**Foreman**	**News**	**Waterfall**	**Football**
Original	2886120	7151874	4115786	6679052	8524180
No. of packets	1961	4859	2796	4537	5791
$\Delta t = 1$	**3124282**	**7365602**	**4498224**	**7237768**	**8617482**
No. of packets	2122	5004	3056	4916	5854
Increase	**8.25%**	**2.98%**	**9.29%**	**8.36%**	**1.09%**
$\Delta t = 2$	**3122252**	**7379988**	**4261158**	**7238698**	**8598686**
No. of packets	2121	5013	2895	4918	5841
Increase	**8.18%**	**3.18%**	**3.53%**	**8.37%**	**0.87%**
$\Delta t = 3$	**3121174**	**7364790**	**4491916**	**7234486**	**8620228**
No. of packets	2120	5003	3052	4915	5856
Increase	**8.14%**	**2.97%**	**9.13%**	**8.31%**	**1.12%**

5.7 Comparison Between IDA, Reed-Solomon and TIRS

In this section, we will first compare TIRS and IDA and then TIRS and Reed-Solomon based approaches, e.g., approaches like the one suggested by Cai and Chen [5].

The compressed size of the video stream will increase for both the IDA and TIRS. The increases of the video size will be different from one video to another. For the IDA, the compressed size of the video increases as the interleaving distance grows. The reason for this is that for long interleaving distances, the frames are moved longer distances, thus reducing the temporal locality. The compression rate using TIRS seems to be rather independent of Δt.

To handle long outages using IDA, we need to increase the interleaving distance. However, even if we use long interleaving distances, two or more consecutive frames can be lost as soon as the length of the outage exceeds the GOP length. For instance, if we use interleaving distance 5 and suffer from an outage of two times the length of the GOP, we will lose two consecutive frames at 9 places (GOP length is 9). Each sequence of two lost frames will be separated by three correct frames (due to using interleaving distance 5). Losing a sequence of two or more consecutive frames will obviously decrease the quality of the video. In TIRS we can handle outages in the rage of Δt to $2\Delta t$, without losing two consecutive frames in the video (see Figure 6 (c) for details).

For short interleaving, e.g., IDA-2 and Δt = 1 second in TIRS, the size increase is more or less the same for IDA and TIRS. This is not so surprising since TIRS is IDA-2 but instead of using the GOP length we use a frame group length of Δt. The only other difference between TIRS and IDA-2 is that in TIRS we reorder the frame groups so that we minimize the number of switches between odd and even frames, thus avoiding unnecessary loss of temporal locality.

For long interleaving that can handle long outages, e.g., IDA-5 and Δt = 3 seconds in TIRS, the size increase is significantly larger in IDA, as the neighbouring frames are separated with a distance of 5 times the GOP length, i.e. 5*9 = 45 frames. In TIRS with Δt = 3 seconds

neighbouring frames are separated with 3 times the transfer rate, i.e. 3*30 = 90 frames. This means that TIRS with Δt = 3 seconds can handle longer outages than IDA-5. In general, the outage length that can be tolerated by IDA is less than that of TIRS, since TIRS can increase the value of Δt without any significant additional overhead.

The number of buffers needed for IDA depends on the interleaving distance (N). For TIRS, we only need 2 buffers, which reduce the implementation complexity. The initial waiting time for the video for IDA is increased as the interleaving distance (N) is increased. For TIRS approach, the initial waiting time is Δt.

An alternative to interleaving approaches like IDA and TIRS is to use Forward Error Correction (FEC). A well-known FEC scheme is Reed-Solomon [4]. Cai and Cheng [5] discuss different ways of interleaving symbols (frames), and thus spreading out the errors in an optimal way before applying Reed-Solomon coding. By spreading out the errors in an optimal way, one gets maximum benefit from the parity symbols. The way that Cai and Cheng use interleaving, however, does not change the fundamental properties of their FEC-based approach. The fundamental difference between the interleaving (such as used in TIRS and IDA) and FEC approaches is that FEC approaches try to transmit all frames correctly by correcting the errors using parity symbols, while the interleaving approaches approximate missing frames by interpolating neighbouring frames. To handle long outages in a FEC-based approach, we either need to spread out the information over very long time intervals, thus generating a significant start-up delay, and/or use a lot of the bandwidth for parity.

For instance, if we want to handle a 1 second outage with a Reed-Solomon code with 10% parity overhead, we need to code the data for the missing second in a 110/5 = 22 seconds block (using the standard notation for Reed-Solomon, we assume that k = 100, t = 5 and a data transmission rate of 5 symbols per second). This means that in order to handle outages up to 1 second, an FEC-based approach would cause a 22 seconds delay instead of a 1 second delay, which we get with TIRS (which has less than 10% overhead).

If we increased the parity overhead in Reed-Solomon to 20%, we would, at least, need a 60/5 = 12 seconds delay for the 1 second outage case (here we assume that k = 50, t = 5 and a data transmission rate of 5 symbols per second). If we increased the parity overhead in Reed-Solomon to 40%, we would need at least 35/5 = 7 seconds delay for the 1 second outage case (here we assume that k = 25, t = 5 and a data transmission rate of 5 symbols per second). In general, we need a $2\Delta t$ $(100 \pm x) / x$ second delay in order to handle a Δt outage using x % parity overhead. This means that there is a very significant cost in terms of overhead and/or delay due to exactly recreating the missing frames with FEC, instead of interpolating them with interleaving approaches such as TIRS and IDA.

5.8 Subjective Viewing Test

5.8.1 Testing Methods

It is well known that the Peak Signal-to-Noise Ratio (PSNR) does not always rank the quality of an image or video sequence in the same way as a human being would. There are many other factors considered by the human visual system and the brain [14].

One of the most reliable ways to assess the quality of a video is subjective evaluation using Mean Opinion Score (MOS). MOS is a subjective quality metric obtained from a panel of human observers. It has for many years been regarded as the most reliable form of user-perceived quality measurements [15].

MOS measurements are used to evaluate the video quality in this study. We 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 range from 0 to 100. These grades are mapped to quality ratings on the 5-grade category scale labelled: Excellent (5), Good (4), Fair (3), Poor (2), and Bad (1). The physical laboratory environment used, with controlled lighting and set-up, conforms to the ITU-R recommendation [10].

The subjective experiment was 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 years.

5.8.2 Testing Materials and Environments

The study is done by coding and compressing the video test sequences Akiyo, Foreman, News, and Waterfall [26] using the H.264 ffmpeg codec [27] for the proposed interleaving technique. The chosen videos are coded with a resolution of 176 x 144. The transmission rate is 30 frames per second and the number of frames transmitted is 1800 (corresponding to 60 seconds).

The outage duration times in this study are 650, 1300 and 2000 milliseconds for short, medium, and long outages respectively. The video sequences are shown on a 17 inch EIZO FlexScan S2201W LCD computer display monitor with a resolution of 1680 x 1050 pixels. The video sequences for the standard technique with no interleaving and TIRS are displayed with a resolution of 176 x 144 pixels in the centre of the screen with a black background.

5.9 Experimental Results

For the TIRS technique, we use Δt = 1 second. This means that we will be able to handle all short outages, some medium outages, and maybe some long outages depending on when the outage occurs. We calculated the conventional statistics such as the average (Avg), the standard deviation (StD), coefficient of variation (CV) and the 95% confidence interval (CI), for the scores. The statistical analysis of the data from the subjective experiments is shown in Figure 7 and Table 3.

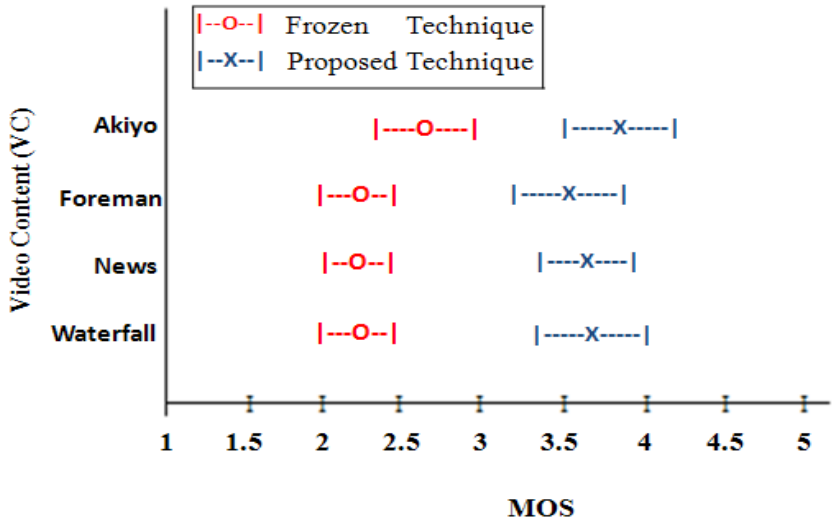
The MOS for TIRS is significantly better than the MOS for the standard technique for all videos. For the standard technique with no interleaving, it is clear that the viewers are disturbed in the frozen pictures in the test videos. For the short outages, the MOS is lower than

3 for all videos, except for the Akiyo video which had a MOS lower than 3.5, on the five-level quality scale. The reason for a higher score for Akiyo is that the video contains less motion than the other videos. For medium and long outages, the MOS is lower than 3, due to the higher percentage of missing frames, which the viewer notices as frozen pictures.

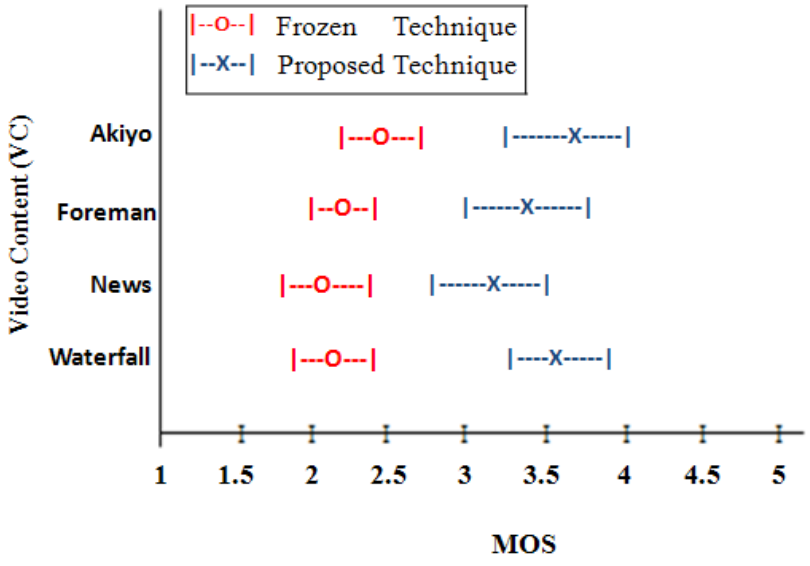
For medium and long outages, the MOS for TIRS is somewhat lower than for the case with short outages. The reason for this is that Δt = 1 second cannot handle all of the medium and long outages. Figure 6 (c) shows that the length of the outage that can be handled depends on when the outage occurs. In our case (Δt = 1 second), we can handle outages of at least 1 second and at best 2 seconds.

Let r denote the vector of the number of ratings in each category provided by the users in a certain experiment, i.e. $r = [r_{\text{Bad}(1)}, r_{\text{Poor}(2)}, r_{\text{Fair}(3)}, r_{\text{Good}(4)}, r_{\text{Excellent}(5)}]$. For each outage duration l, we obtain one vector of ratings r^l without taking advantage of interleaving, and a corresponding vector $r^{l,\text{TIRS}}$ when using TIRS. The difference vector $\Delta r = r^{l,\text{TIRS}} - r^l$ 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 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 ($\Delta r_{\text{Good}(4)} \geq 0$, $\Delta r_{\text{Excellent}(5)} \geq 0$). Figure 8 supports this assumption and also 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.

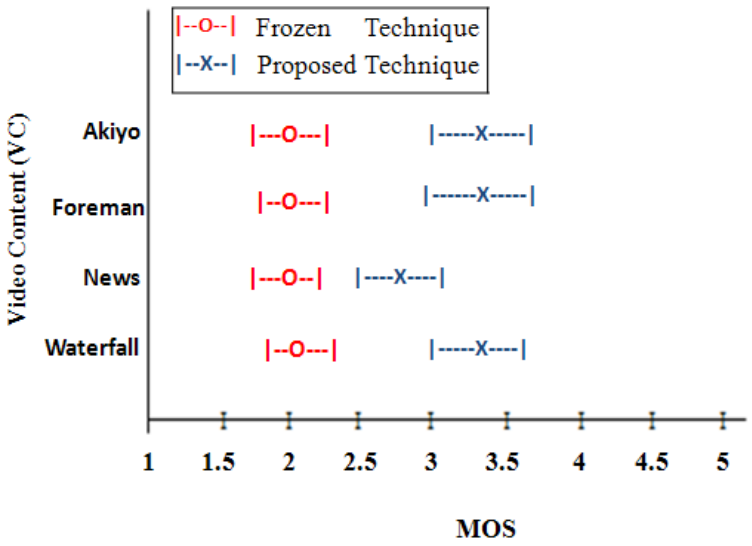
The videos presented to the viewers resulted in a wide range of perceptual quality ratings in the experiments. When we analyse the scores we feel that TIRS is a satisfactory technique to distribute the missing frames in the streaming video, to eliminate the frozen pictures and provide a smooth video on the mobile screen.



a. MOS for short outage time



b. MOS for medium outage time



c. MOS for long outage time

Figure 7: The MOS for different videos contents and for different outage time, showing the average and the standard deviation.

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

a. Statistical data analysis for short outage time

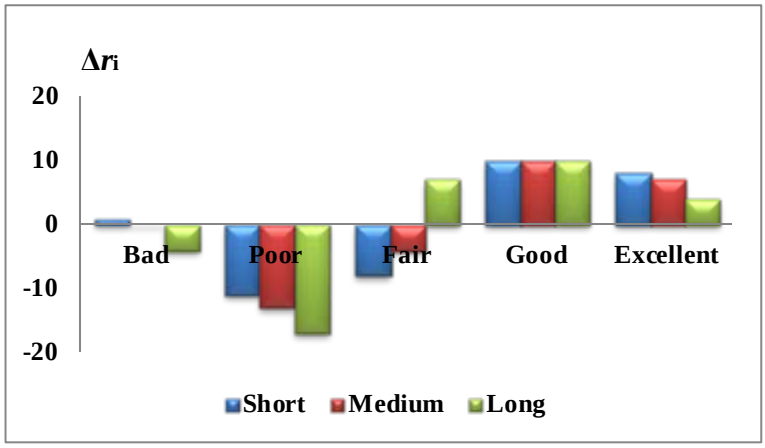
	Frozen Technique				Proposed Technique			
	Akiyo	**Foreman**	**News**	**Waterfall**	**Akiyo**	**Foreman**	**News**	**Waterfall**
Avg	2,70	2,27	2,27	2,27	3,87	3,57	3,67	3,70
StD	0,65	0,52	0,45	0,52	0,97	0,90	0,66	0,92
CoV	24%	23%	20%	23%	25%	25%	18%	25%
95 % CI	9,0%	8,5%	7,4%	8,5%	9,3%	9,4%	6,7%	9,3%

b. Statistical data analysis for medium outage time

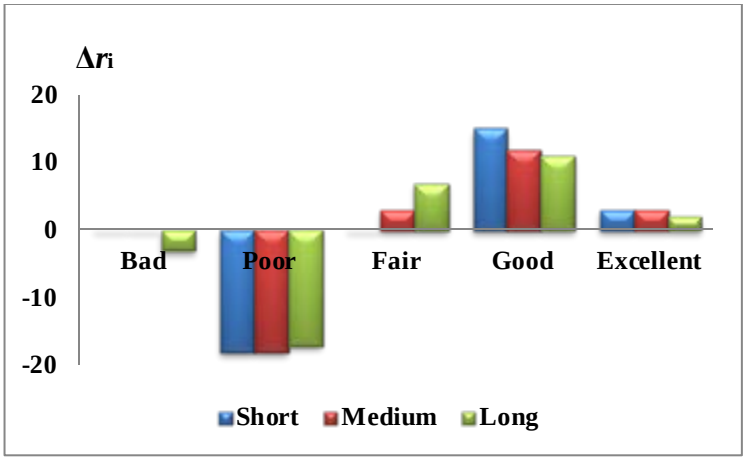
	Frozen Technique				Proposed Technique			
	Akiyo	**Foreman**	**News**	**Waterfall**	**Akiyo**	**Foreman**	**News**	**Waterfall**
Avg	2,40	2,20	2,03	2,17	3,63	3,40	3,13	3,63
StD	0,56	0,48	0,61	0,53	1,07	0,97	0,97	0,96
CoV	23%	22%	30%	24%	29%	29%	31%	26%
95 % CI	8,7%	8,1%	11,2%	9,1%	11,0%	10,6%	11,5%	9,8%

c. Statistical data analysis for long outage time

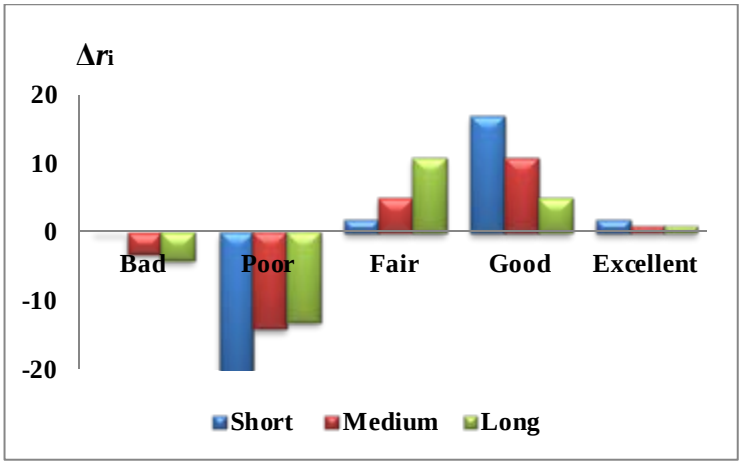
	Frozen Technique				Proposed Technique			
	Akiyo	**Foreman**	**News**	**Waterfall**	**Akiyo**	**Foreman**	**News**	**Waterfall**
Avg	2,00	2,03	1,93	2,10	3,43	3,30	2,87	3,37
StD	0,59	0,56	0,58	0,55	0,97	0,92	0,90	0,96
CoV	30%	28%	30%	26%	28%	28%	31%	28%
95 % CI	11,0%	10,3%	11,2%	9,8%	10,5%	10,4%	11,7%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



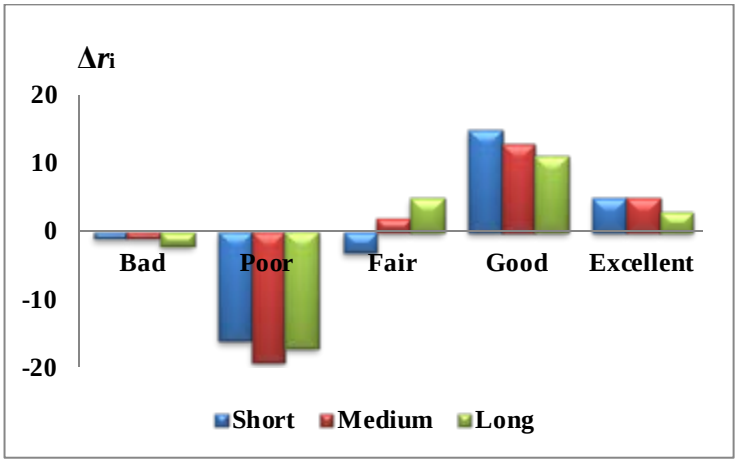
a. Akiyo video



b. Foreman video



c. News video



d. Waterfall video

Figure 9: Difference in the number of user ratio per category when employing the TIRS technique and for different outage times.

5.10 Conclusion

Wireless network transmission errors in forms of frame losses could have a major impact on the end user experience of real time videos. A Time Interleaving Robust Streaming (TIRS) technique is proposed to minimize the effect of outages and frame losses. Streaming the video frames according to TIRS makes it possible to significantly reduce negative effects when a sequence of consecutive video frames is lost during the transmission.

The advantage of TIRS is that, the lost frame sequences will be spread out on the streaming video with the ability to reconstruct it at the receiver side. This means that if 30 consecutive frames are lost, we do not get a 1 second freezing (assuming 30 frames per second); the 30 frames are distributed to 2 frame groups related to another 2 frame groups. Because of this, the lost frames 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%).

TIRS offers some important advantages compared to the previous IDA interleaving technique when we have long outages (more than 1 second). The first advantage is that the nature of the interleaving scheme in IDA is such that even if we use long interleaving distances, two or more consecutive frames can be lost as soon as the length of the outage exceeds the (Group of Pictures) GOP length. In TIRS, we can handle outages in the range of Δt to $2\Delta t$, without losing two consecutive frames in the video as shown in Figure 6 (c). The second advantage of TIRS compared to IDA is that, in TIRS the interleaving length (i.e. Δt) can be increased without affecting the size overhead. In IDA, the size overhead increases significantly when the interleaving distance grows.

Our evaluation was based on user panel tests and showed that there is a significant quality improvement when using TIRS.

References

[1] 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.

[2] 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.

[3] 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).

[4] Benslimane A., 2007. Multimedia Multicast on the Internet. Wiley-ISTE publisher.

[5] 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.

[6] 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.

[7] 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).

[8] 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).

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

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

[25] 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.

[26] trace.eas.asu.edu/yuv/index.html.(visited, 1/11/2009)

[27] www.ffmpeg.org. (visited, 4/03/2011)

CHAPTER SIX

Compressing Video Based on Region of Interest

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

Abstract

Real-time video streaming suffer from bandwidth limitation that are unable to handle the high amount of video data. To reduce the amount of data to be streamed, we propose an adaptive technique to crop the important part of the video frames, and drop the part that are outside the important part; this part is called the Region Of Interest (ROI). The Sum of Absolute Differences (SAD) is computed to the consecutive video frames on the server side to identify and extract the ROI. The ROI are extracted from the frames that are between reference frames based on three scenarios. The scenarios been designed to position the reference frames in the video frames sequence. Linear interpolation is performed from the reference frames to reconstruct the part that are outside the ROI on the mobile side. We evaluate the proposed approach for the three scenarios by looking at the size of the compressed videos and measure the quality of the videos by using the Mean Opinion Score (MOS). The results show that our technique significantly reduces the amount of data to be streamed over wireless networks with acceptable video quality are provided to the mobile viewers.

Keywords

Sum of Absolute Differences, Region of Interest, Reference Frames, Video Compression, Mean Opinion Score

6.1 Introduction

Bandwidth is one of the most critical resources in wireless networks, and the available bandwidth of wireless networks should be managed efficiently [10]. Therefore, the size of a video stream should be adapted according to the network bandwidth [5],[6]. Network adaptation refers to how much network resources (e.g., bandwidth) a video stream should be utilize i.e., it is an adaptive streaming mechanism for video transmission [4].

Video Coding [8], like H.264, is developed to encode/decode the video frames with respect to the specific rate required by a certain application. In a low bit-rate video coding, such as on mobile video streaming, the video could skip some temporal/spatial levels to meet the bandwidth limitations.

The main feature of H.264/SVC [11] is to provide bandwidth-optimized transmission for video streaming by observing current network conditions. H.264/SVC provides three types of enhancements for optimized bandwidth transmission. First, it can support temporal enhancements by changing the frame rate. Second, it can support spatial enhancements through resolution, and finally it can support enhancements of the quality through a signal-noise-rate.

The basic element of H.264/AVC video sequence is slicing, where each frame can be divided into several slices [7] and each slice contains a group of macroblocks (MBs) [3],[12]. H.264/AVC introduces Flexible Macroblock Ordering (FMO) as a useful error resilient tool, where H.264/AVC defines seven different types of FMO modes: Interleaved, Dispersed, Fore-ground with left-over, Box-out, Raster-scan, Wipe, and Explicit [16]. The H.264/SVC encodes the video in the way that can be selectively transmitted according to the type option; contents and network condition by using a bit stream extractor [11].

In this paper, we present a video adaptation technique to reduce the amount of data to be streamed over wireless network. The streaming server will identify and extract the high motion slice region (ROI) from the frames that are between reference frames and drop the less motion slice region (non-ROI). Four different ROIs for three different

scenarios are proposed to study the effect of the compression size on the video streaming.

After the mobile device has received the video stream, linear interpolation between reference frames will be performed to reconstruct the pixels that are outside the ROI (non-ROI). Mean Opinion Score (MOS) measurement is used to evaluate the quality of the reconstructed videos.

6.2 Background and Related Work

Several techniques have been proposed for spatial adaptation for slicing the video frames. Mavlankar et al. [2] examine how to determine the slice sizes for streaming the ROI. In their work, the server will adapt and stream according to the regions size of the video content that is desired at the client's side. They study the trade-off in the choice of slice size. The optimal slice size achieves the best trade-off to minimize the expected number of bits that are transmitted to the client per frame. The output of their work is to predict the optimal slice size regions, which depends on the signal as well as the display resolution for the mobile screen.

Moiron et al. [15], proposed a slicing scheme for enhanced error resiliency. The proposed scheme used FMO without introducing any increases in the bit rate or computational complexity. The frame level priority is provided based on the slice position within the frame. The frame was sliced into three distinct regions. The slice structuring in the encoder was modified to accommodate a new set of rules for video slicing. These rules prevent macroblocks from different regions from being packed into the same slice. They also define when the current slice should be terminated.

Wang et al. [17] applied a cropping technique to perform spatial adaptation to the video stream to overcome the display constraints by producing the ROI and according to the mobile screen resolution. The ROI is determined by using an attention based modelling method. The ROI is automatically detected and crop the informative region in each frame to generate a smooth video sequence.

6.3 The Proposed Technique

In the related work section, the researchers identify the ROI as the most attractive region to the viewers. The ROI is extracted from the video frames to cope with the bandwidth limitation and the mobile screen resolution. In this study, the Sum of Absolute Differences (SAD) metric is used to identify the motion region position which we considered it as the ROI. The ROI will be extracted from the video frames in the server and stream it over the network. On the receiver side, the pixels that are outside the ROI (non-ROI) will be reconstructed from the reference frames by using linear interpolation [18].

6.3.1 Identifying the ROI

The SAD technique is a commonly used technique for motion estimation in various video encoding standards like H.264 [1]. The idea is to take the absolute differences between consecutive video frames. The SAD value will be zero except for the changes induced by the objects moving between the video frames. If there is a lot of motion in one region of the frames, the SAD value in this region will be relatively high, and if there is no motion then the SAD value in this region will be zero. The SAD is computed to identify the position of the ROI. The idea is to place the ROI where there is a lot of motion in the video scene and drop the less motion region (non-ROI). We assume that we could have different ROIs for different sequences in a video.

Two ways of scanning the consecutive video frames are considered to identify the highest intra-slice differences, as shown in Figure 1 and Figure 2, respectively:

- The first way is to scan the consecutive video frames from top-to-bottom, as shown in Figure 1.

$$SAD_H(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sum_{x=0}^{L-1} |F_x(i, j+k) - F_{x-1}(i, j+k)| \qquad (1)$$

- The second way is to scan the consecutive video frames from left-to-right, as shown in Figure 2.

$$SAD_V(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sum_{x=0}^{L-1} |F_x(i+k, j) - F_{x-1}(i+k, j)| \qquad (2)$$

where L is the length of the frame sequences, N x M, is the height and width for the ROI, and k is a fixed region.

The test videos are used in this work were the samples of video sequences Highway, Akiyo, Foreman, News, and Waterfall, with a resolution of 176 x 144 pixels [19]. The chosen videos are well known as professional test videos that have different characteristics.

Figure 1: Scanning the slice region based on $SAD_H(k)$.

Figure 2: Scanning the slice region based on $SAD_V(k)$.

For Highway video, the SAD_H value is the largest when the ROI is close to the bottom of the frames as the highest differences among the intra-slices as shown in Figure 3(a). The SAD_V value for Highway video is the highest when the ROI is close to the sides of the frames as shown in Figure 3(b).

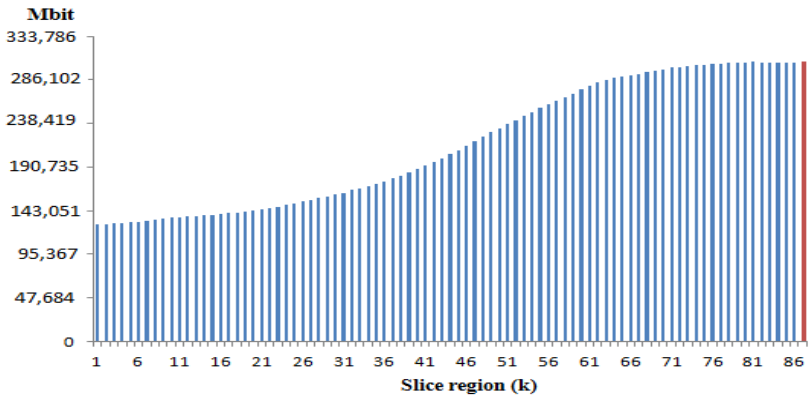
For Akiyo and News videos, the SAD_H and SAD_V values are the highest in the middle of the frames as shown in Figure 4 and Figure 6, respectively.

For Foreman video, the SAD_H value is the highest at the bottom of the frames and the SAD_V value is the highest in the middle of the frames as shown in Figure 5. It is been notice from that, the SAD_H and SAD_V values are approximately within the same range as the Foreman video is shaking all the times.

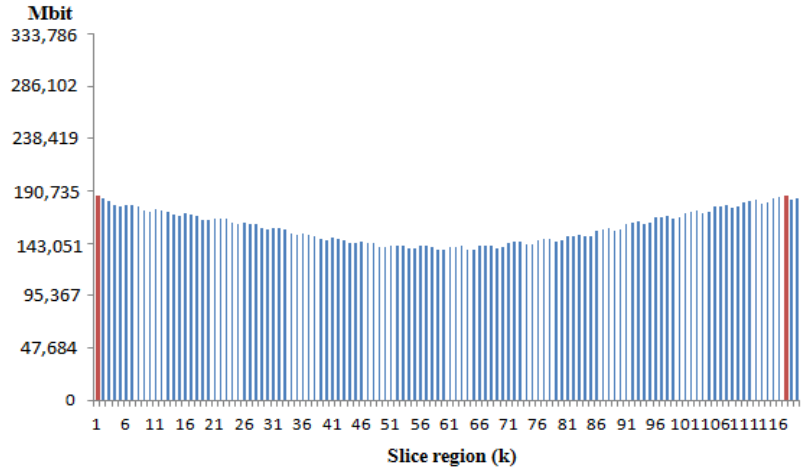
For Waterfall video, the SAD_H value is increasing dramatically as the video is zooming out all the times, where the highest value is in the bottom of the frame, while the SAD_V value it is the highest on the left side of the frames as shown in Figure 7.

From the result that we are obtained from calculating the SAD values; four different ROIs cases are proposed in this study as shown in Figure 8:

- ROI case A corresponds to the highest value of the SAD_H; see Figure 3(a), Figure 4(a), Figure 5(a), Figure 6(a) and Figure 7(a).
- ROI case B corresponds to the highest value of the SAD_V; see Figure 3(b), Figure 4(b), Figure 5(b), Figure 6(b) and Figure 7(b).
- ROI case C is chosen statically in the middle of the frames as shown in Figure 8 (c) and Table 1.
- ROI case D is actually two regions. The reason for investigating this ROI is that some video, e.g., Highway (see Figure 3(b)), has a lot of motion on the right and left side of the video frames as shown in Figure 8 (d) and Table 1.

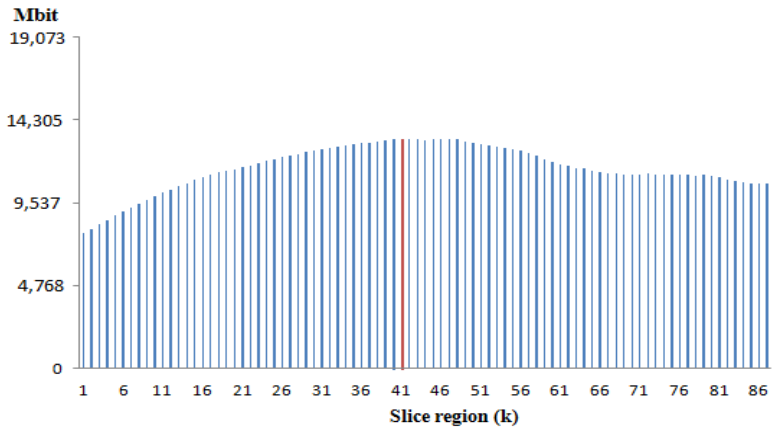


a. The SADH (*k*)

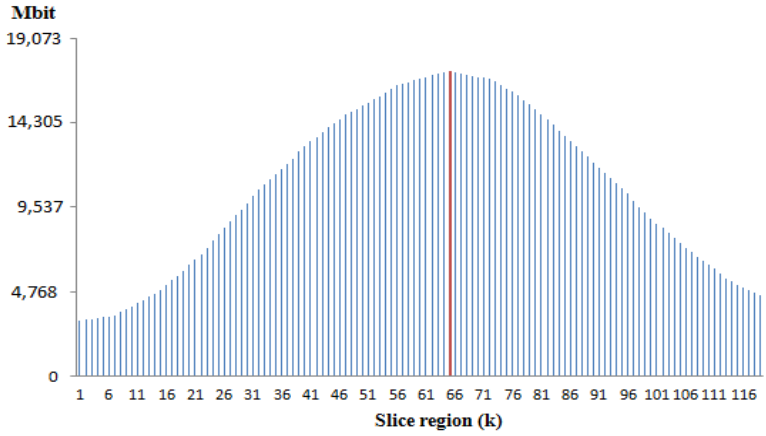


b. The SADV (*k*)

Figure 3: Highway video.

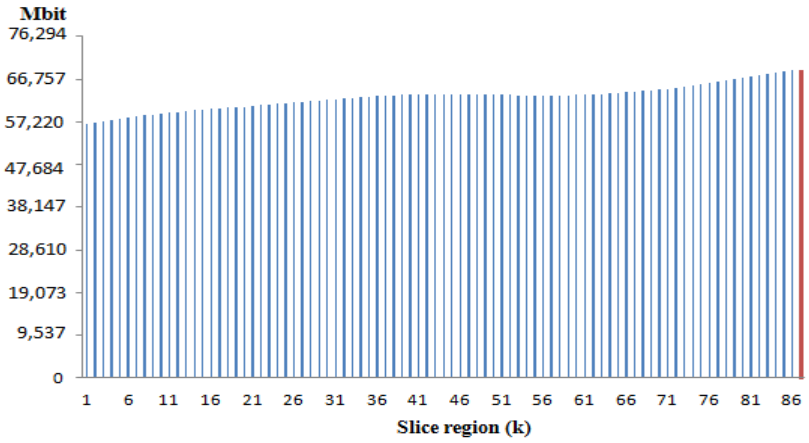


a. The SADH (*k*)

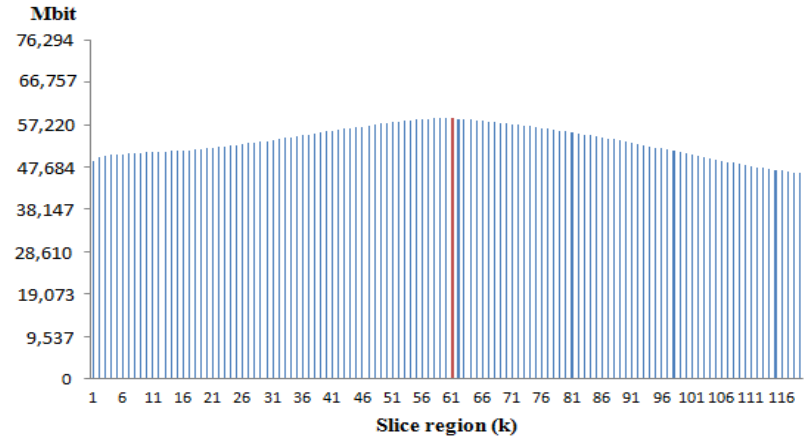


b. The SADV (*k*)

Figure 4: Akiyo video.

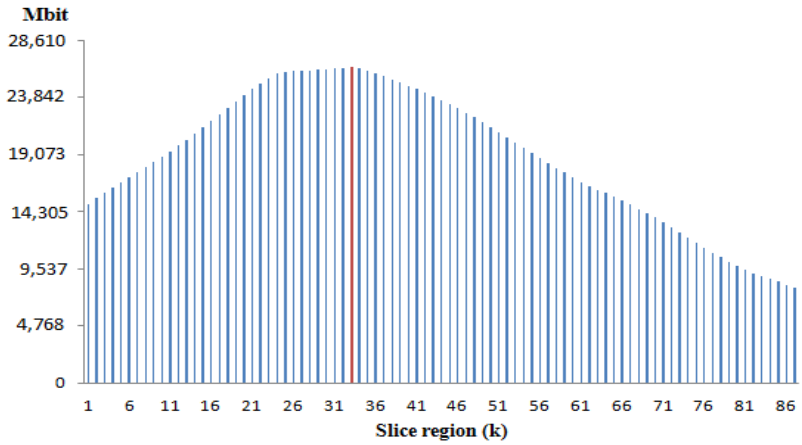


a. The SADH (*k*)

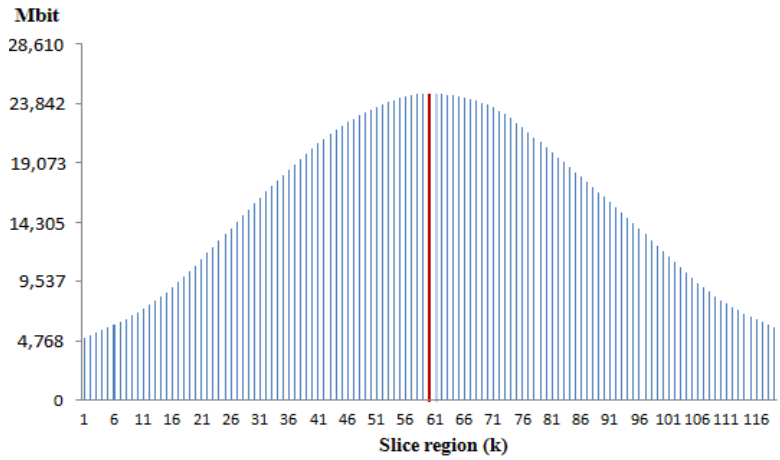


b. The SADV (*k*)

Figure 5: Foreman video.

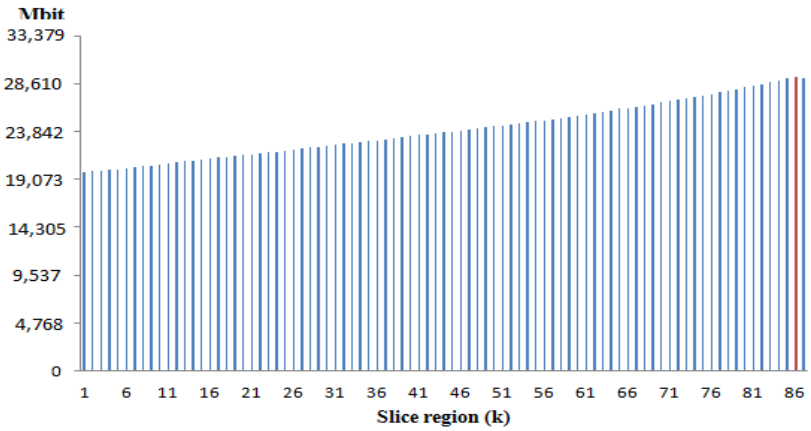


a. The SAD_H (k)

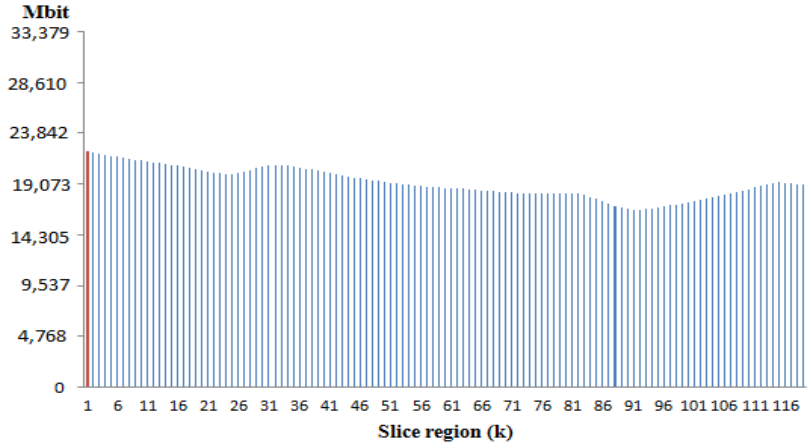


b. The SAD_V (k)

Figure 6: News video.



a. The SADH (*k*)



b. The SADV (*k*)

Figure 7: Waterfall video.

176 X 58 = 10208 pixels

a. Case A

58 X 144 = 8352 pixels

b. Case B

101 X 101 = 10201 pixels

c. Case C

(36 + 36) X 144 = 10368 pixels

d. Case D

Figure 8: The proposed ROI.

a. The reference frames are every 3rd frame

b. The reference frames are every 4th frame

c. The reference frames are every 5th frame

Figure 9: The proposed scenarios.

Table 1: The SAD values (Mbit) for case C and D.

Videos	Highway	Akiyo	Foreman	News	Waterfall
Case C	2018.591	17.789	67.018	29.901	22.718
Case D	247.313	3.636	60.173	4.391	27.739

6.3.2 The Proposed Streaming Scenarios

The streaming video is divided into two videos: one video containing a complete version of every n^{th} frames in the original video sequence, which we call it the reference frames and one video containing the ROI that are between reference frames. In this study we consider three scenarios of n (n = 3, 4, and 5), as shown in Figure 9.

In the first scenario where the reference frames are every 3^{rd} frame, e.g., 0, 3, 6, 9,..., in the second scenario where the reference frames are every 4^{th} frame, e.g., 0, 4, 8, 12,..., in the third scenario where the reference frames are every 5^{th} frame, e.g., 0, 5, 10, 15,...

The two videos (one with every n^{th} frames and one with the ROI) are encoded and transmitted in the normal way by using H.264 [20].

At the mobile side, the receiving video frames are been reordered to it is original sequence position, while linear interpolation is performed between reference frames to reconstruct the part of the region that is outside the ROI (non-ROI). This means that the value of a pixel (x,y) outside of the ROI for frames $k*n+i$ (k = 0, 1, 2, 3… and i = 0,…, n-1) is obtained by using the following formula $((n-i)Fkn\ (x,y) + iF(k+1)n\ (x,y))/n$, where $F_{kn}(x,y)$ is the pixel value for position (x,y) in frame number kn in the original video sequence.

6.4 The Effect of the Proposed Technique on the Streaming Size

Compressing the video frames according to the proposed technique by using H.264 ffmpeg codec [20] will affect the amount of redundancy information to be removed from the video frames. Since we will drop the pixels that are outside the ROI (non-ROI), we expect that the compression size of the videos will be affected.

The proposed scenarios are applied for different cases (ROIs) to the videos Highway, Akiyo, Foreman, News, and Waterfall [19], with a coding rate of 30 frames per second. The chosen videos have different characteristics and motions level. Therefore, the compression size of

the videos is expected to be different from one video to another and from one scenario to another.

The compression size of the videos is decreases when the proposed ROI method is applied, as the amount of decrease is different from one scenario to another and from one ROI to another. For Highway videos, ROI case A shows the smallest decrease as shown in Table 2 (a). The reason for that, the ROI case A has the highest SAD value and the non-ROI that are related to the ROI case A is more static compared to the non-ROI that are dropped from the other ROIs, as shown in Figure 3 and Table 1.

Table 2: The compressed video size (bytes).

a. Highway

	Case A	**Case B**	**Case C**	**Case D**
Scenario one	25891300	23470760	24326420	25241910
No. of packets	17589	15945	16526	17148
Decrease	**34.5%**	**40.6%**	**38.5%**	**36.1%**
Scenario two	24626600	21835484	22809872	23834016
No. of packets	16730	14834	15496	16196
Decrease	**37.7%**	**44.8%**	**42.3%**	**39.7%**
Scenario three	23685770	20741686	21772332	22887146
No. of packets	16091	14091	14791	15548
Decrease	**40.1%**	**47.5%**	**44.9%**	**42.1%**

b. Akiyo

	Case A	**Case B**	**Case C**	**Case D**
Scenario one	12679446	15669696	15418140	9685038
No. of packets	8614	10645	10474	6580
Decrease	**30.7%**	**14.3%**	**15.7%**	**47.0%**
Scenario two	11739910	15044034	14787634	8414166
No. of packets	7975	10220	10046	5716
Decrease	**35.8%**	**17.8%**	**19.2%**	**54.0%**
Scenario three	11120176	14569282	14286604	7600286
No. of packets	7554	9898	9706	5163
Decrease	**39.2%**	**20.3%**	**21.9%**	**58.4%**

c. Foreman

	Case A	Case B	Case C	Case D
Scenario one	29306084	29182400	29005722	29542048
No. of packets	19909	19825	19705	20069
Decrease	**36.6%**	**36.9%**	**37.3%**	**36.1%**
Scenario two	26949156	26750004	26565964	27182634
No. of packets	18308	18173	18046	18466
Decrease	**41.7%**	**42.2%**	**42.6%**	**41.2%**
Scenario three	25331482	25165654	24951476	25598484
No. of packets	17209	17096	16951	17390
Decrease	**45.2%**	**45.6%**	**46.0%**	**44.6%**

d. News

	Case A	Case B	Case C	Case D
Scenario one	19146848	19093548	20964438	15158016
No. of packets	13007	12971	14242	10298
Decrease	**26.8 %**	**27.0%**	**19.8%**	**42.0%**
Scenario two	17809400	17799214	19882266	13443716
No. of packets	12099	12092	13507	9133
Decrease	**31.9%**	**31.9%**	**24.0%**	**48.6%**
Scenario three	16762434	16574180	18756994	12303162
No. of packets	11388	11260	12743	8358
Decrease	**35.9%**	**36.6%**	**28.3%**	**52.9%**

e. Waterfall

	Case A	Case B	Case C	Case D
Scenario one	30666804	29842876	30307382	31064252
No. of packets	20833	20274	20589	21103
Decrease	**34.6%**	**36.3%**	**35.3%**	**33.7%**
Scenario two	28110148	27215516	27722092	28525000
No. of packets	19097	18489	18833	19378
Decrease	**40.0%**	**41.9%**	**40.8%**	**39.1%**
Scenario three	26546070	25609170	26142780	26972772
No. of packets	18034	17398	17760	18324
Decrease	**43.3%**	**45.3%**	**44.2%**	**42.4%**

For Akiyo videos, the compression size is decreases, as shown in Table 2 (b). The ROI for cases B and C shows the lowest decreases. These are also the ROIs with the largest SAD value, as shown in Figure 4 and Table 1.

For Foreman videos, the decreases of the compression size for all ROIs are approximately within the same range as the video is shaking all the time, as shown in Table 2 (c). The SAD value is also approximately within the same range for all cases as shown in Figure 5 and Table 1.

For News videos, the ROI case C shows the lowest decreases in the compression size, as shown in Table 2 (d). This is also the ROI with the largest SAD value, as shown in Figure 6 and Table 1.

For Waterfall videos, the amount of compression size is within the same range for all ROIs, since the video is zooming out. However, ROI case D has somewhat lower compression size as shown in Table 2 (e). Looking at Figure 7 and Table 1, we see that the SAD value for ROI case D is the highest.

It can be summarized from the compression Tables, the size of the compressed videos are related to the SAD value to the corresponding ROI.

6.5 Subjective Viewing Test

6.5.1 Testing Materials and Environments

The videos are displayed on a 17 inch FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of 1680 x 1050 pixels. The videos are displayed with resolution of 176 x 144 pixels in the centre of the screen with a black background with a duration of 66 seconds for Highway video and 10 seconds for Akiyo, Foreman, News and Waterfall videos.

6.5.2 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 are considered by the human visual system and the brain [13]. One of the most reliable ways of assessing the quality of a video is subjective evaluation of the Mean Opinion Score (MOS) measurement, 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 [14]. The MOS measurements are used to evaluate the videos quality in this study and based on the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). We use a lab with controlled lighting and set-up according to the ITU-R recommendation. The score grades in this methods range from 0 to 100. These ratings are mapped to a 5-grade discrete category scale labelled with Excellent, Good, Fair, Poor and Bad [9].

The subjective experiment has been conducted at Blekinge Institute of Technology in Sweden. The users observed the proposed scenarios in two groups, where the participated of thirty non-expert test subjects for each group were all university students. The participated of the first group are 25 males and 5 females for evaluating the Highway videos and their age's range of 20 to 41. The participated for the second group are 27 males and 3 females for evaluating Akiyo, Foreman, News, and Waterfall videos and their age's range of 20 to 35.

The amount of data is gathered from the subjective experiments with respect to the opinion scores that were given by the individual viewers. Concise representation of this data is achieved by calculating the conventional statistics such as the mean score and 95% confidence interval.

6.6 Experimental Results

For Highway videos, the first scenario (every 3rd frame), as shown in Figure 9 (a) and Figure 10 (a). The observers evaluate the videos after

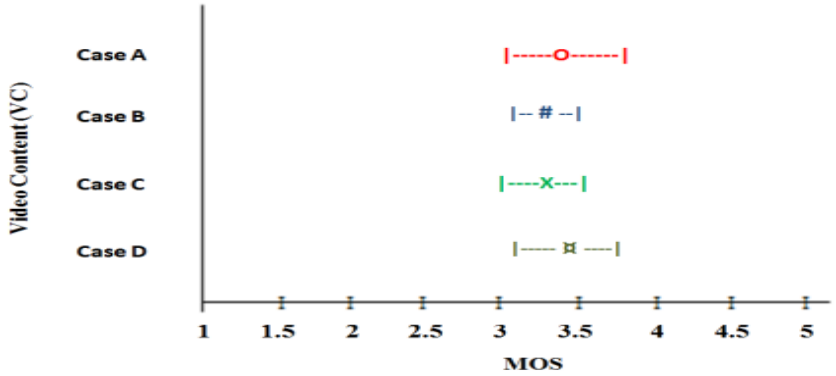
it been reconstructed by using linear interpolation. The scores are within the same range for the four ROI cases, as the MOS is larger than 3 and lower than 4. However, the MOS for ROI case A is slightly higher than the rest. The second scenario (every 4^{th} frame), as shown in Figure 9 (b) and Figure 10 (b). The MOS vary from one case of ROI to another and the ROI case D shows the lowest score, while the ROI case A shows the better score than the rest. For the third scenario (every 5^{th} frame), as shown in Figure 9 (c) and Figure 10 (c), the MOS for ROI case A shows the highest score as ROI case A has the highest SAD value as well, as shown in Figure 3 and Table 1.

The MOS for Akiyo videos as shown in Figure 11. The ROI case D shows the lowest score for the three scenarios. From Table I, the SAD value for ROI case D is significantly smaller than other ROIs. As the ROI case D is been defined as a two side regions as shown in Figure 8 (d), where the highest motion region is in the middle of the frames as shown in Figure 4. Therefore, after the videos have been reconstructed by interpolation, the motion region is been highly affected and for this reason the MOS is low for ROI case D. The differences in both MOS and SAD value for ROIs A, B and C is rather small for Akiyo video.

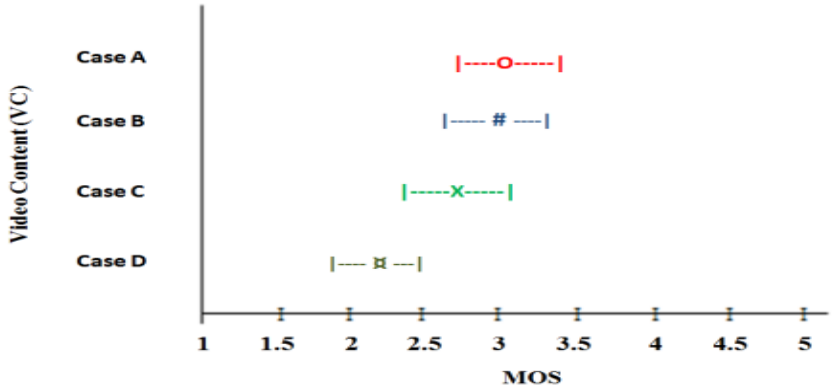
The MOS for Foreman videos as shown in Figure 12. The scores are below 2.5 for the four ROI cases, as the Foreman video frames are shaking all the time, which is negatively been affected by the interpolation.

The MOS for News videos as shown in Figure 13. The Figure shows that the ROI case C gets the highest MOS, as the highest motion is in the middle of the frames, while ROI case D gets the lowest MOS, as the motion region been affected by interpolation. From Figures 6, 8 and Table I, we observed from that the ROI case C has the highest SAD value and ROI case D has the lowest SAD value.

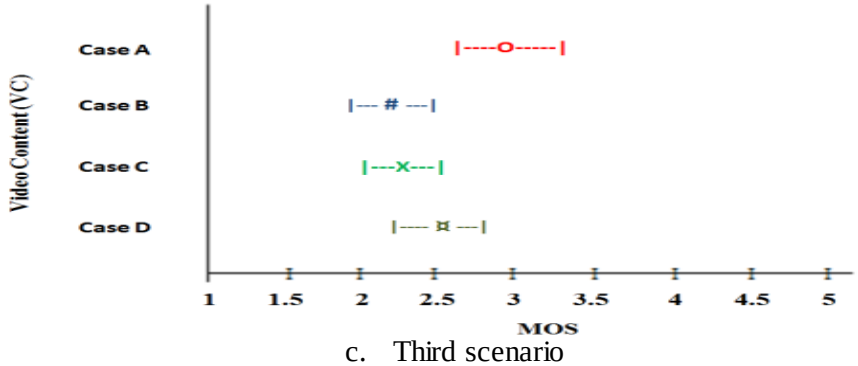
The MOS for Waterfall videos as shown in Figure 14. The scores are greater than 2.5 for the four ROI cases. ROI case D gets the lowest MOS but ROI actually has the highest SAD value, even all the ROIs have approximately similar SADs values. This is an exception to what we have been seen in other videos.



a. First scenario

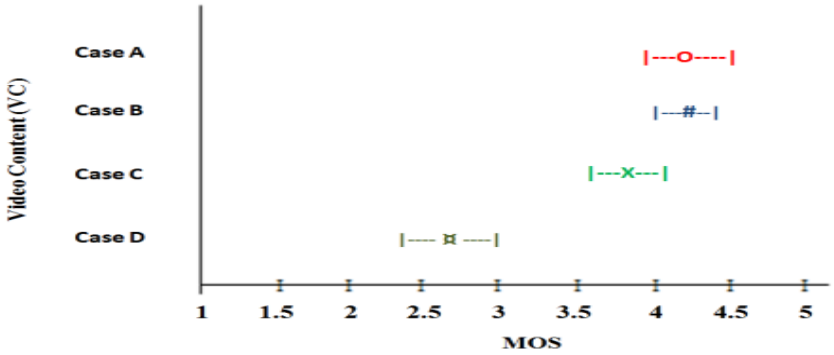


b. Second scenario

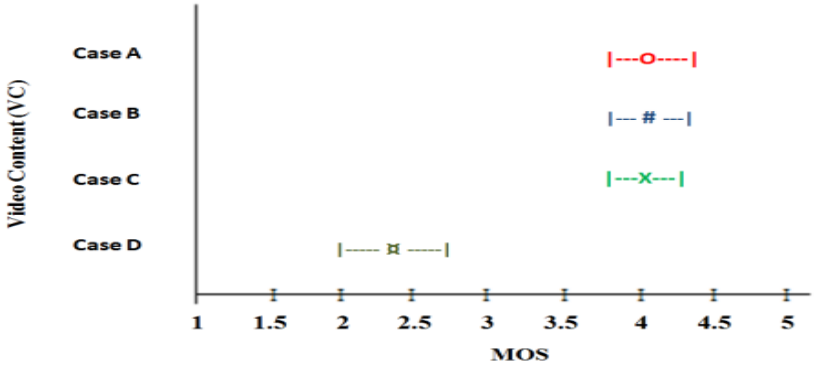


c. Third scenario

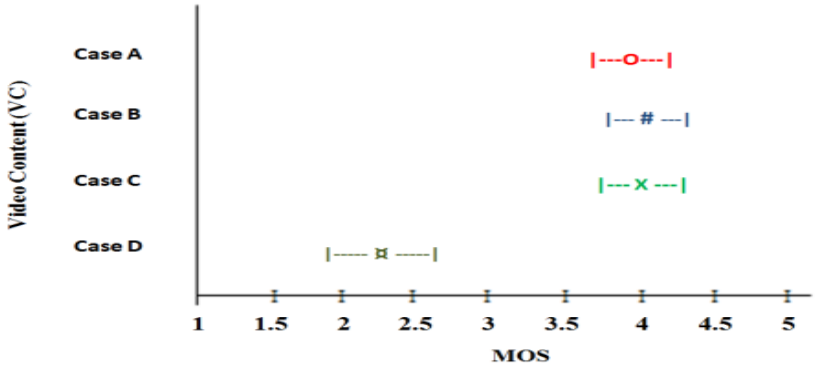
Figure 10: The MOS for the four proposed slice cases for Highway.



a. First scenario

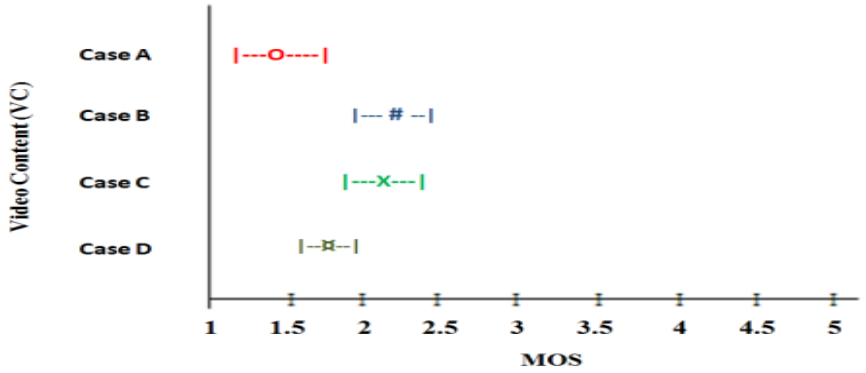


b. Second scenario

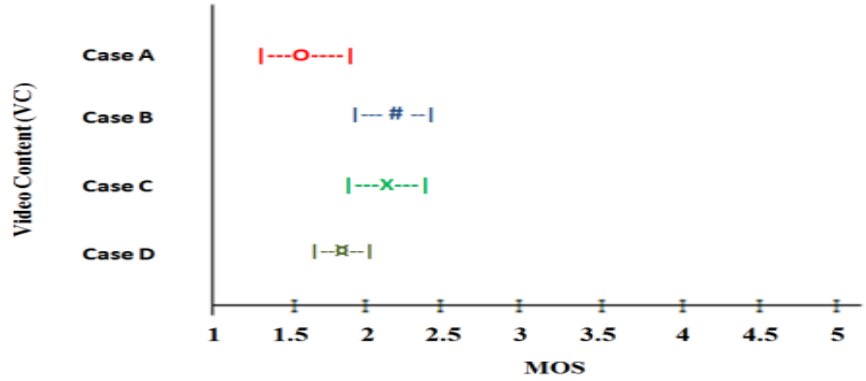


c. Third scenario

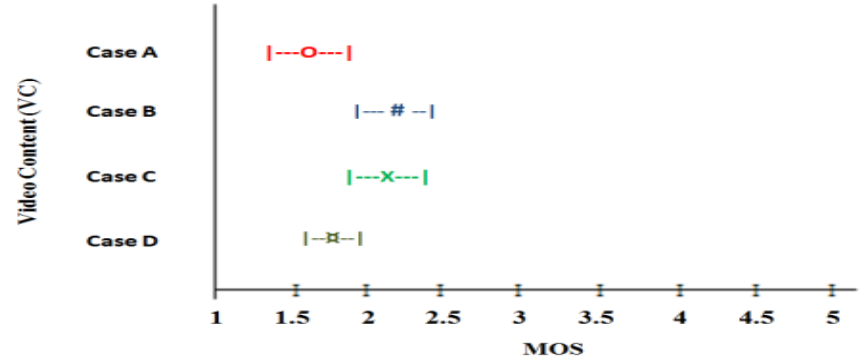
Figure 11: The MOS for the four proposed slice cases for Akiyo.



a. First scenario

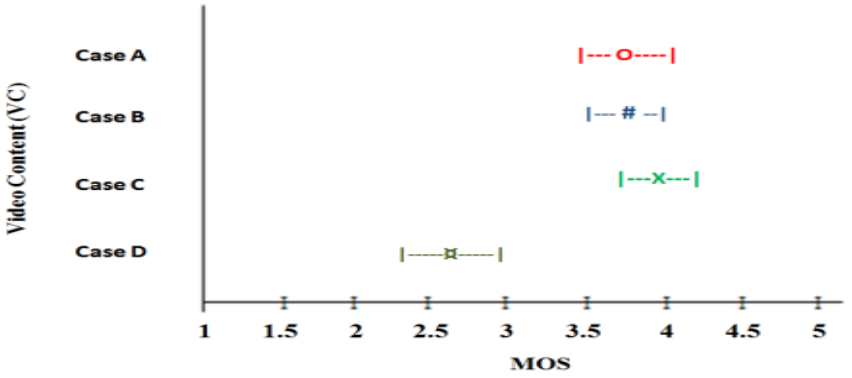


b. Second scenario

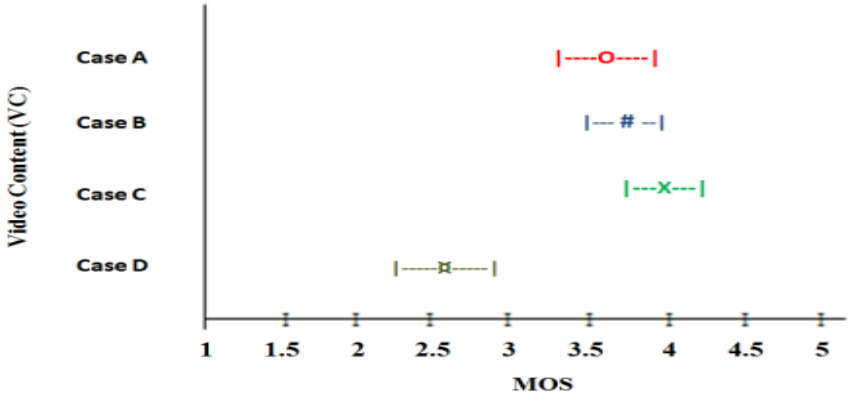


c. Third scenario

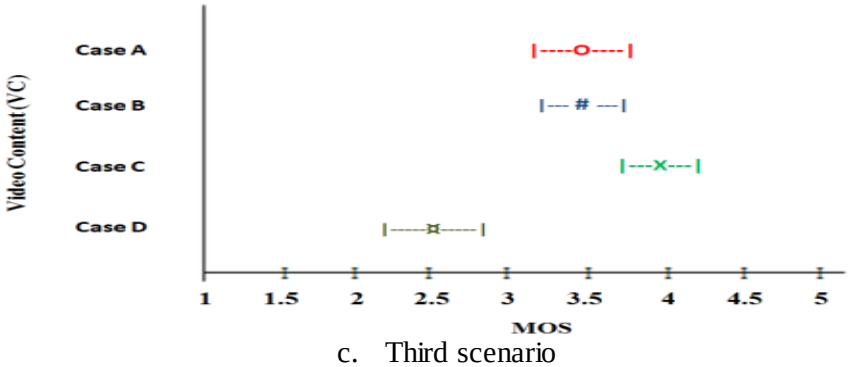
Figure 12: The MOS for the four proposed slice cases for Foreman.



a. First scenario

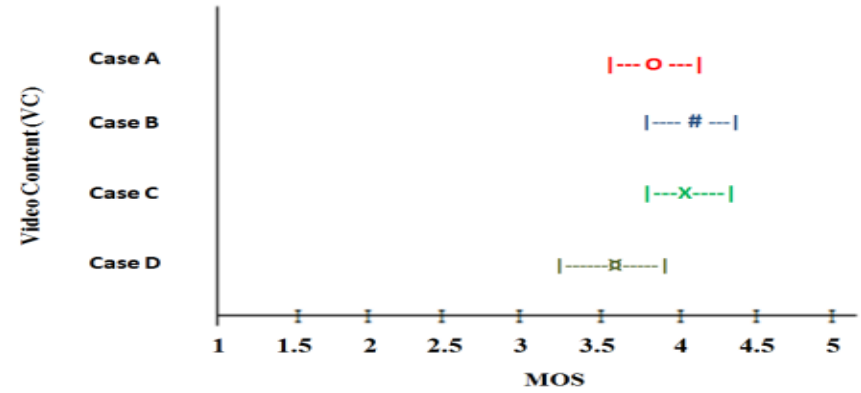


b. Second scenario

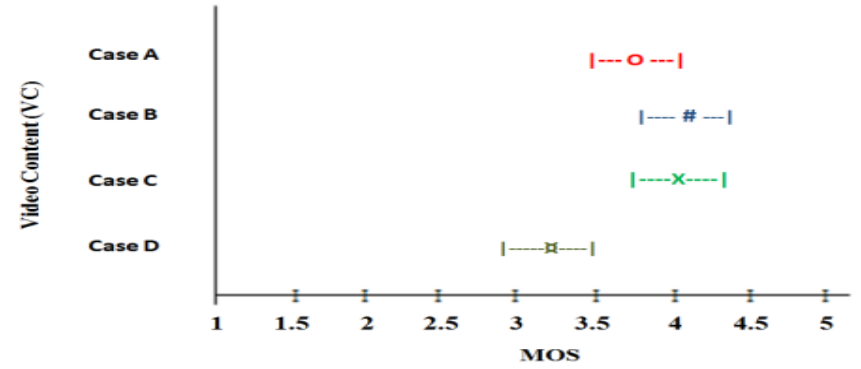


c. Third scenario

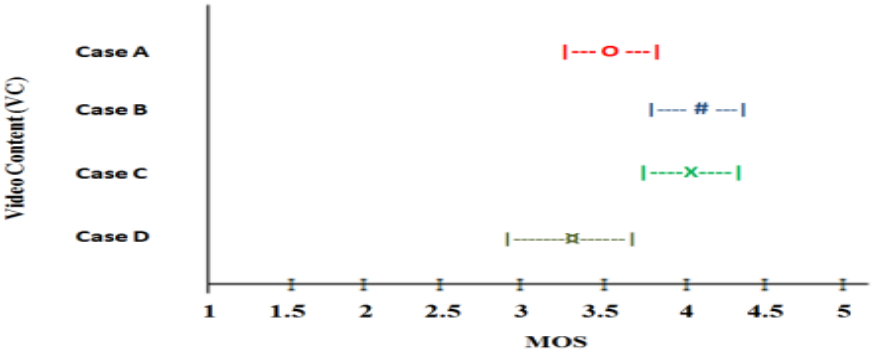
Figure 13: The MOS for the four proposed slice cases for News.



a. First scenario



b. Second scenario



c. Third scenario

Figure 14: The MOS for the four proposed slice cases for Waterfall.

In general, it seems that a high value of SAD results in a high MOS. However, it seems that if the SAD value does not differ so much between different ROIs, e.g., Foreman and Waterfall videos, then ROIs of case B and C seem to be promising.

Figures 10 - 14 shows that by sending the reference frames more often, e.g., every 3^{rd} frame instead of every 4^{th} and 5^{th} frame will increase the MOS in most cases.

6.7 Conclusion

To reduce the amount of data that are streamed over wireless networks, a Region Of Interest (ROI) compression scheme is proposed in this study. The scheme is based on identifying and extracting the motion ROI by computing the Sum of Absolute Differences (SAD) for video streaming.

Four different ROI cases for three different scenarios are proposed to evaluated each video, corresponding to send the full frame information (reference frames) every 3^{rd}, 4^{th}, and 5^{th} frame, respectively.

The proposed adaption scheme is compressed by H.264 codec to study the effects on the video size, by sending the reference frames more often, e.g., every 3^{rd} frame instead of every 4^{th} and 5^{th} frame; it will increase both the MOS and the file size.

Our experiment shows, and quantifies, the trade-off between high compression and high MOS. How often we will send the reference frames is an engineering decision that depends on the available bandwidth and the need for compression level. However, since it is probably more important to keep the MOS on a high level, we would argue that one should select ROIs with the highest SADs values.

The SAD can easily been calculated for different parts (scenes) of a video. We could then use this for obtaining an appropriate ROI for that particular part of the video, i.e., we do not need to use the same ROI for the entire video.

REFERENCES

[1] A. Dimou, O. Nemethova, and M. Rupp, "Scene change detection for H.264 using dynamic threshold techniques," in Proc. EURASIP'05, 2005.

[2] A. Mavlankar, P. Baccichet, D. Varodayan, and B. Girod, "Optimal slice size for streaming regions of high resolution video with virtual Pan/Tilt/Zoom functionality," in Proc. EUSIPCO'07, 2007, p.1275.

[3] D. Grois, E. Kaminsky, and O. Hadar, "ROI adaptive scalable video coding for limited bandwidth wireless networks," in Proc. WD'10, 2010, p. 1.

[4] F. Yang, Q. Zhang, W. Zhu, and Y-Q. Zhang, "Bit allocation for scalable video streaming over mobile wireless internet," in Proc. INFOCOM'04, 2004, p. 2142.

[5] 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," in Proc. CIT'06, 2006.

[6] H. Luo, M-L. Shyu, and S-C. Chen, "An end-to-end video transmission framework with efficient bandwidth utilization," in Proc. ICME'04, 2004, p. 623.

[7] H. Liu, W. Zhang, S. Yu, and X. Yang, "Channel-aware frame dropping for cellular video streaming," in Proc. ICASSP'06, 2006, p. 409.

[8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, pp. 1103-1120, Sept. 2007.

[9] (2002) International Telecommunication Union. Methodology for the Subjective Assessment of the Quality of Television Pictures. ITU-R, Rec. BT.500-11 website. [Online]. Available: http://www.itu.int/rec/R-REC-BT.500-11-200206-S/en.

[10] 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, pp. 1566 – 1574, Dec. 2003.

[11] J-H. Lee, and C. Yoo, "Scalable ROI algorithm for H.264/SVC-based video streaming," IEEE Transactions on Consumer Electronics, vol. 57, pp. 882-887, May, 2011.

[12] L. Al-Jobouri, M. Fleury, S.S.Al-Majeed, and M. Ghanbari, "Effective video transport over WiMAX with data partitioning and rateless coding," in Proc. CIT'10, 2010, p. 767.

[13] 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. MV'07, 2007 p. 31.

[14] M. Martinez-Rach, O. López, P. Piñol, M. P. Malumbres, J. Oliver, and C. T. Calafate, "Behaviour of quality assessment metrics under packet losses on wireless networks," in Proc. XIX Jornadas de Paralelismo, 2008.

[15] S. Moiron, I. Ali, M. Ghanbari, and M. Fleury, "Enhanced slicing for robust video transmission," in Proc. EUROCON'11, 2011, P. 1.

[16] S. Gong, J. Yang, and S. Zhang, "A Novel content-based video coding scheme for robust video transmission over Ad Hoc networks," in Proc. WiCom '09, 2009, P. 1.

[17] Y. Wang, X. Fan, H. Li, Z. Liu, and M. Li, "An attention based spatial adaptation scheme for H.264 videos on mobiles," in Proc. MMMC'06, 2006.

[18] Y-C. Peng, H-A. Chang, C-K. L. Chen, and H. C-J. Kao, "Integration of image stabilizer with video codec for digital video cameras," in Proc. ISCAS'05, 2005, p. 4871.

[19] Video Test Sequence. [Online]. Available: trace.eas.asu.edu/yuv/index.html.

[20] Video codec. [Online]. Available: www.ffmpeg.org.

CHAPTER SEVEN

Adapting the Streaming Video based on the Estimated Motion Position

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

Abstract

In real-time video streaming, the frames must meet their timing constraints, typically specified as their deadlines. Wireless networks may suffer from bandwidth limitations. To reduce the data transmission over the wireless networks, we propose an adaption technique on the server side by extracting a part of the video frames that is considered as a Region Of Interest (ROI), and drop the part outside the ROI from the frames that are between reference frames. The estimated position of the selection of the ROI is computed by using the Sum of Squared Differences (SSD) between consecutive frames. The reconstruction mechanism to the region outside the ROI is implemented on the mobile side by using linear interpolation between reference frames. We evaluate the proposed approach by using Mean Opinion Score (MOS) measurement. MOS are used to evaluate two scenarios with equivalent encoding size, where the users observe the first scenario with low bit rate for the original videos, while for the second scenario the users observe our proposed approach. The results show that our technique significantly reduces the amount of data that are streamed over wireless networks, while the reconstruction mechanism provides acceptable video quality.

Keywords

Streaming Video, Region of Interest, Sum of Squared Differences, Mean Opinion Score

7.1 Introduction

Nowadays mobile cellular networks are able to support different type of services, such as video streaming that makes a great demand on wireless networks bandwidth. Bandwidth is the most critical resource in mobile networks [1]; therefore, it is important to employ adaption mechanism for efficient use of the available bandwidth. Network adaptation refers to how much network resources (e.g., bandwidth) a video stream should be utilize for video content, resulting in designing an adaptive streaming mechanism for video transmission [2].

The main feature of H.264/SVC is to provide bandwidth-optimized transmission for real time video streaming by observing the current network conditions [3]. H.264 contains a rate-control algorithm that are dynamically adjusts the encoder parameters to achieve a target bit rate by allocates a budget of bits to the video frames sequence. The main concept of rate-control algorithm is a quantitative model, which describes the relationship between the quantization parameter and the actual bit rate [4].

The quantization parameter (QP) has a great impact on the encoder performance, because it regulates on how much spatial details can be saved. As the increases of the QP, some of the details are aggregated so that the bit rate drops with some increases in distortion and some losses of the video quality [5]. The frame size can be reduced to eliminate the artifacts at low bit rate environment. However, the reduction of the size does not guarantee a good quality, as the original video contents in high resolution, where the video quality will be poor when the bit rate is low [6].

The limitation of the available bit rate is one of the key technologies that required efficiently allocating bits for the purpose of video contents for transmitting the Region Of Interest (ROI) [7]. The user attention is the ability to detect the interest parts for a given scene that called attention area or ROI [8]. The ROI can be extracted from the streaming video, as the ROI consider the most interesting and important parts in the video frame, while the background (non-ROI) are dropped as it is considered less important region.

The major idea to encode the ROI in the video is to reduce the bit rate by sacrificing the quality of the non-ROI; the other is to allocate more bandwidth to enhance the quality of the ROI or important factor for determination the quantization parameters (QP) in the encoder [7].

In this paper, we present an adapting technique to reduce the amount of data to be streamed over the wireless networks. The streaming server will set the reference frames and extract the slice region from the frames that are between reference frames. After the mobile device received the video stream, linear interpolation between references frames are performed to reconstruct the pixels that has been dropped on the server side. Mean Opinion Score (MOS) measurement that are obtained from a panel of human will observe and evaluate the videos after the dropping pixels (non-ROI) are reconstructed.

7.2 Related Work

Several techniques have been proposed for spatial adaptation for slicing the video frames. Wang and El-Maleh [9] proposed an adaptive background (non-ROI) skipping approach where every two consecutive frames are grouped into a unit. In each unit, the first non-ROI will perform encoding, while the second non-ROI is skipped (using predicted macroblocks with zero motion vectors). The ROI are either identifies automatically or been specified by the end-user, while the non-ROI will be skipped and the numbers of bits are allocated to the ROI is to ensure the best visual quality of the video.

Shuxi et al. [10] proposed a spatial domain adjustable resolution method based on the ROI. The proposed method is to divide the video frame into ROI and non-ROI. The ROI have more details, as it is the most important region in the video frames, while the details of the non-ROI are ignored. The ROI will perform encoding on the same resolution of the original frame. The non-ROI will perform coding on the low resolution. Coding the frames from the region of non-interest after down-sampling in order to achieve adjustable resolution feature based-on the region. They claimed that the proposed method could reduce the complexity of the encoding to guarantee the subjective and the objective quality of the ROI.

Inoue et al. [8] proposed data format based on Multiview Video Coding (MVC) for two types of partial delivery method with and without lower-resolution. The two types of partial delivery method are considered in their work for multi-bit rates and resolution to maximize the partial panoramic video quality under restricted bandwidth. The first type is partial delivery method that is used to deliver the frames without lower-resolution; while the second partial delivery method with lower-resolution. In their work, they examined the impact of subject image quality in terms of delivery method and the ROI movement. The work examined three delivery methods, 'Deliver-all', 'Partial delivery' with and without lower-resolution. They claimed that the two types of the proposed partial delivery methods could achieve higher subjective video quality than the deliver-all methods when the ROI are on the move.

The above researchers identify the ROI as the most attractive object or region to the viewers. Some researcher considered two different resolutions for encoding the video frame, a high resolution for the ROI and low resolution for the non-ROI. Others researchers considering skipping the non-ROI to provide a good quality to the ROI that can cope with the bandwidth limitations.

In our study, we define the ROI as the most motion region within the video frames that are calculated from the sum of the motion differences between the video frames, while the less motion region are considered as the non-ROI.

7.3 The Proposed Technique

The Sum of Squared Differences (SSD) metric is computed to detect the most motion region, which we call it the ROI. The ROI will be extracted from the frames between reference frames on the server sides, and drop the pixels outside ROI. On the mobile side, the part of the region that is outside the ROI (non-ROI) will be reconstructed by using linear interpolation between reference frames, as shown in Figure 1.

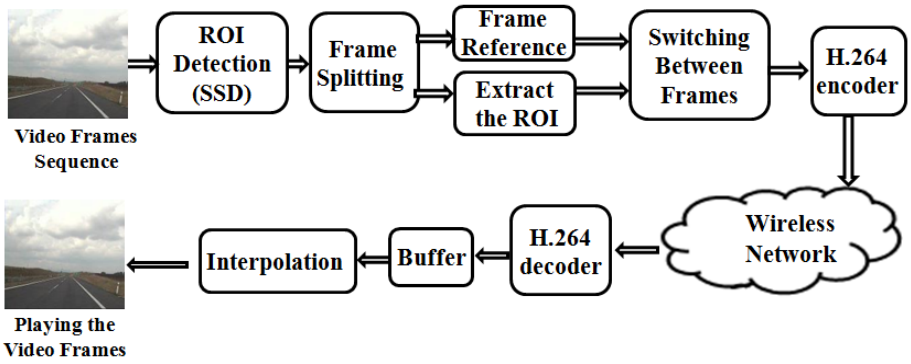


Figure 1: The proposed streaming technique.

7.3.1 Detecting the ROI

The SSD technique is a commonly used technique for motion estimation for video codec standard like H.264 [11]. Computing the SSD for the consecutive video frames will be similar except for the changes that might be induced by the objects moving within the frames.

Figure 2: Scanning the slice region based on SSD (*k*).

The SSD is computed to detect and identify the estimated motion position of the slice by scanning the consecutive video frames from top to bottom based on the highest intra-slice differences. The highest SSD value is an indicator that it is the most motion region in the video frames that are considered as the ROI, as shown in Figure 2 and according to 1:

$$\mathrm{SSD}\ (k) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sum_{x=0}^{L-1} \left(F_x(i, j+k) - F_{x-1}(i, j+k)\right)^2 \quad (1)$$

where L is the length of the frame sequences, N x M, is the height and width, while k is a fixed region.

The test videos are used in this work were the samples of video sequences Highway, Akiyo, Foreman, News, and Waterfall, with a resolution of 144 x 176 [12]. The videos been chosen as they have different characteristics.

For Highway video and as shown in Figure 3. The SSD value is the lowest in the top of the frames as an indicator that, there is less activities, therefore the motion is less. The SSD value is the highest in the bottom of the frames as there are high activities in the video frames, therefore the motion is high.

For Waterfall video and as shown in Figure 4. The SSD value is increasing dramatically from the top of the frames until the bottom of the frames, as the video is zooming out all the times where the SSD value is the highest in the bottom of the frames as an indicator that, it is the most motion region.

For News video and as shown in Figure 5. The SSD value is the lowest in the top and in the bottom of the frames but it is slighter higher in the middle part that is closed to the top of the frames as more activates been detected, therefore the motion is high.

For Foreman video and as shown in Figure 6. The SSD value it is approximately within the same range, because the Foreman video is

shaking all the times, therefore, the SSD value is relatively high in all regions but it is slighter higher in the bottom of the frames.

For Akiyo video and as shown in Figure 7. The SSD value is the lowest in the top and in the bottom of the frames, while it is the highest in the middle of the frames as an indicator that there are high activities, therefore the motion is high.

The highest value that we are obtain from computing the SSD for the consecutive video frames is the most motion and important region, therefore it is considered as the ROI in this work.

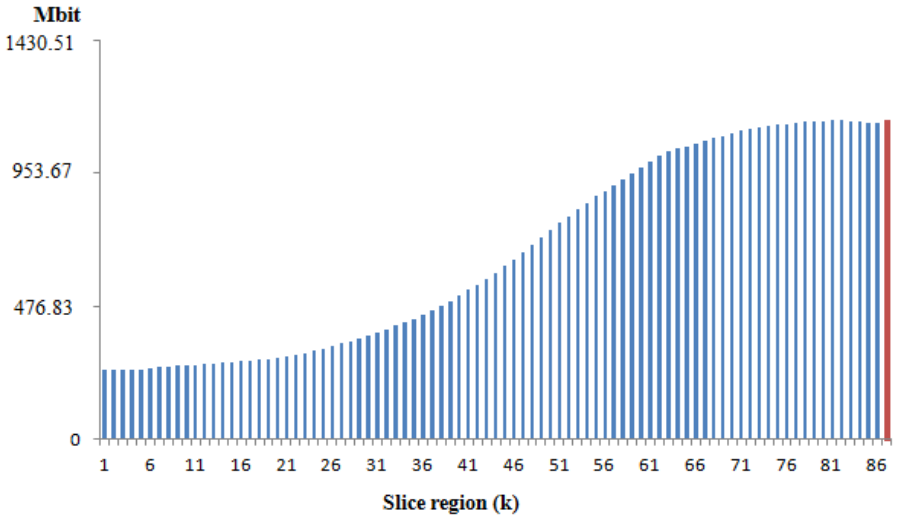


Figure 3: The SSD (*k*) for Highway video.

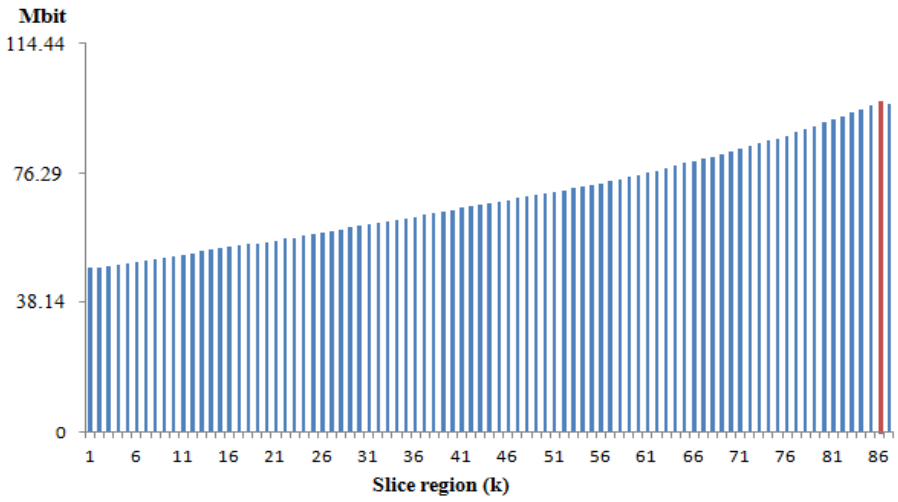


Figure 4: The SSD (k) for Waterfall video.

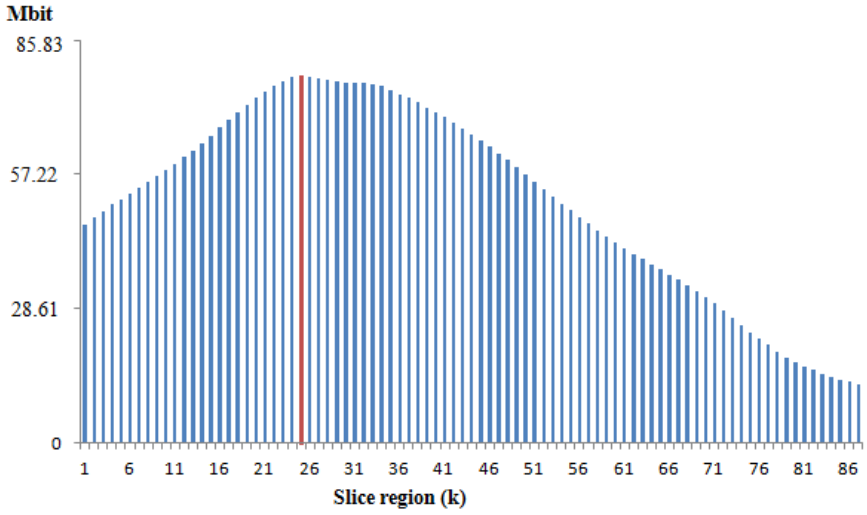


Figure 5: The SSD (k) for News video.

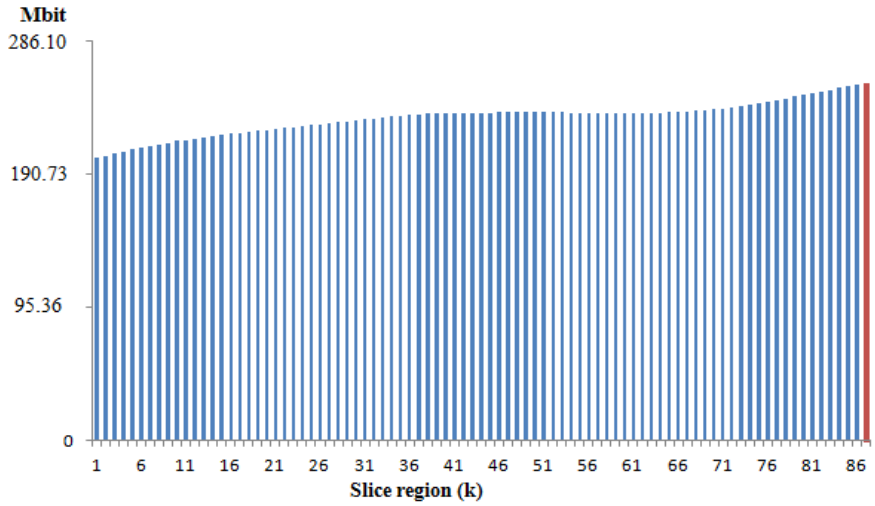


Figure 6: The SSD (k) for Foreman video.

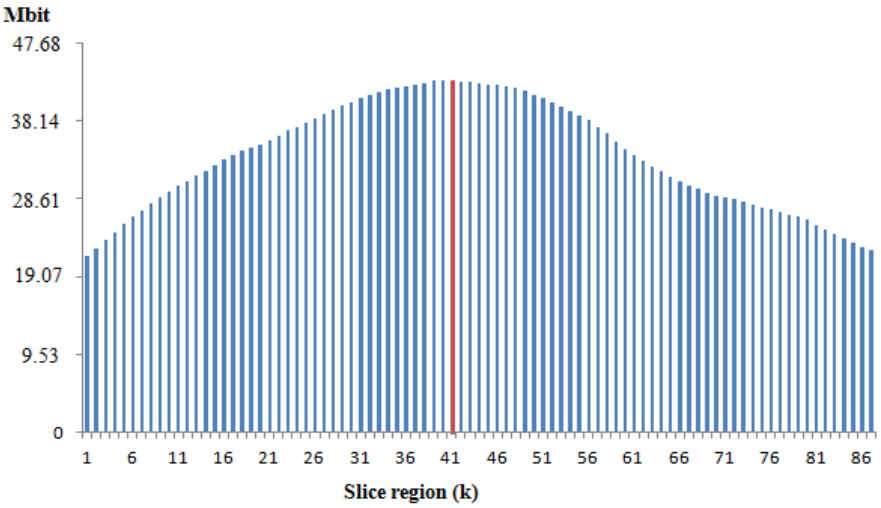


Figure 7: The SSD (k) for Akiyo video.

7.3.2 Extracting the ROI

The streaming server will establish the connection according to the mobile request. The server will compute the SSD value to the consecutive video frames, to detect and extract the position of the ROI based on the highest intra-slice differences and drop the pixels that are outside the ROI.

The sequences of the video frames are split to set the reference frames and to extract the ROI from the frames between reference frames. The sequence position of the reference frames are considered in this study is every fifth frame, as the maximum distance between frames that do not have high effect on the quality of the viewers' perception [13].

The reference frames with the ROI are combined by a switching mechanism for encoding and transmitting in the normal way by using H.264 codec, as shown in Figure 1.

7.3.3 Reconstructing the Video Frames

After the mobile device start receiving the video frames (reference frames and ROI), it will be held in the buffer to reconstruct the surrounding pixels that are outside the ROI (non-ROI) that been dropped on the server.

The method is used to reconstruct the non-ROI is linear interpolation [14]. Linear interpolation is applied between reference frames to reconstruct the non-ROI pixels. After the frames been returned to their original shape, the video will be played on the mobile screen.

7.4 Quantization Parameter Adaptation

The video will encode to obtain the optimum visual quality within the available networks bandwidth. The bit allocation for the video should achieve the tradeoff between encoding video quality and bandwidth limitation.

The bit allocation for the video is encoded by using H.264 ffmpeg codec [15]. The videos are encoded to identify the effectiveness of the bit rates and the quantization parameters (QP) on the encoding size. The encoding size will be different from one video to another as the videos had different characteristics.

Two scenarios are proposed to encode the videos, the first scenario; where the original videos are encoded with a default QP for a bit rate of 64 kbps. The second scenario (the proposed scenario), where the videos are encoded with a bit rate of 128 kbps, as shown in Table 1. The main idea to encode the videos in the second scenario with adaptive QP is to gain equivalent encoding size for the videos that are in the first scenario that can cope with the limitation of the network bandwidth.

Table 1: The encoding videos size for the two scenarios.

Test Videos	Size (KB), QP=2	Scenario 1: QP	Size (KB)	Coding Efficiency Gain	Scenario 2: QP	Size (KB)	Coding Efficiency Gain
Highway	4564	14	707	**84.50%**	10	693	**84.81%**
Waterfall	1280	10	165	**87.10%**	8	175	**86.32%**
News	1128	11	167	**85.19%**	7	172	**84.75%**
Foreman	616	12	146	**76.29%**	9	139	**77.43%**
Akiyo	350	6	165	**52.85%**	4	172	**50.85%**

7.5 Subjective Viewing Test

7.5.1 Test Methods

It is well known that the Peak Signal-to-Noise Ratio (PSNR) does not always rank the 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 [16].

One of the most reliable ways of assessing the quality of a video is subjective evaluation of 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 [17].

7.5.2 Test Materials and Environments

The videos are displayed on a 17-inch FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of 1680 x 1050 pixels. The videos are displayed with resolution of 176 x 144 pixels in the center of the screen with a black background with a duration of 66 seconds for Highway video and 10 seconds for Akiyo, Foreman, News and Waterfall videos.

The MOS measurement are used to evaluate the videos quality in this study and based on the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). We use a lab with controlled lighting and set-up according to the ITU-R recommendation. The score grades in this methods range from 0 to 100. These ratings are mapped to a 5-grade discrete category scale labelled with Excellent, Good, Fair, Poor and Bad [18].

The subjective experiment was conducted at Blekinge Institute of Technology in Sweden. The participant of thirty non-expert test subjects, 25 males and 5 females. They were all university students and their ages range from 20 to 35. The users observed two scenarios for displaying the videos; the first scenario is to display the videos for a

low bit rate for the original videos and the second scenario by implementing our proposed technique with linear interpolation for a high bit rate. The playing rates for both scenarios are 30 frames per second.

The amount of data are gathered from the subjective experiments with respect to the opinion scores that were been given by the individual viewers. Concise representation of this data are achieved by calculating the conventional statistics such as the mean score and 95% confidence interval [18].

7.6 Experimental Results

A panel of users evaluates the two scenarios according to the Mean Opinion Score (MOS) measurement. In the first scenario, the original videos are decoded with a bit rate of 64 kbps. The second scenario (the proposed scenario), where the observers evaluates the videos after been decoded with a bit rate of 128 kbps and performed linear interpolation to reconstruct the pixels that are outside the ROI, as shown in Figure 8.

For Highway videos, the observers evaluate both scenarios within the same score range, as an indicator that the observers had similar opinion to the quality of the videos.

For Waterfall videos, the MOS score for both scenarios is larger than 3 and less than 4, while the score for the first scenario is slightly higher than the second scenario.

For News and Akiyo videos, the MOS for the second scenario had better score than the first scenario, as the ROI fit in the motion region as shown in Figure 5 and Figure 7, respectively. Therefore, the observers did not manage to recognize the effect of interpolation on the video frames.

For Foreman videos, the score of the MOS for the second scenario is the lowest score than the first scenario, as the observers manage to recognize the effect of interpolation on the video frames, although the first scenario is been coded with a low bit rate.

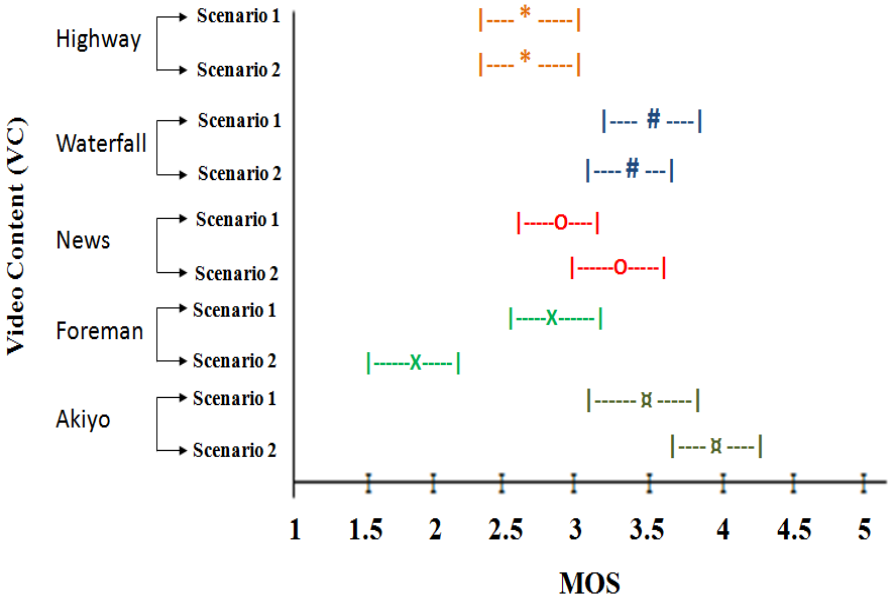


Figure 8: The MOS for different videos content.

7.7 Conclusion

In this study, we proposed an adaptive scheme to identify and extract the appropriate slice (ROI) by computing the Sum of Squared Differences (SSD) on the server side and drop the pixels that are outside the ROI. The receiving video on the mobile device will reconstruct the dropping pixels that are outside the ROI by using linear interpolation and from the reference frames.

In general, it seems that the highest SSD results as an indicator to the important region in the video frames. A panel of users observers and evaluates the two scenarios by using the MOS measurements. The user's panel observed the first scenario with a low bit rate for the original videos and the second scenario with a high bit rate (the proposed adaptive scheme for estimate the position of ROI).

It is been notice from that, the MOS score is the highest for the videos like Waterfall, Akiyo and News, while for a video like Foreman, the MOS is the lowest score as it is not easily to estimate the ROI as the video frames are shaking all the time.

Even the quality of the videos is degraded; it could still be a satisfactory technique for reducing the encoding size of the streaming video over limited bandwidth.

References

[1] CHANG, Jau-Yang and CHEN, Hsing-Lung. Service-Oriented Bandwidth Borrowing Scheme for Mobile Multimedia Wireless Networks. Proceedings International Conference on Wireless Broadband and Ultra Wideband Communications (Auswireless 2006).

[2] SCHWARZ, Heiko, MARPE, Detlev, WIEGAND, Thomas. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Transactions on Circuits and Systems for Video Technology [online]. July 2007, vol. 17, iss. 9, [cit. 2007-09-24]. ISSN 1051-8215. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4317636.

[3] LEE, Jung-Hwan and YOO, Chuck. Scalable ROI Algorithm for H.264/SVC-Based Video Streaming. Proceedings IEEE Transactions on Consumer Electronics. Las Vegas: 2011. p. 201-202. ISSN : 2158-3994.

[4] CZUNI, Laszlo, CSASZAR, Gergely and LICSAR, Attila. Estimating the Optimal Quantization Parameter in H.264. Proceedings 18th International Conference on Pattern Recognition (ICPR). Hong Kong: 2006. p. 330-333. ISBN: 0-7695-2521-0.

[5] HRARTI, M., SAADANE, H., LARABI, M., TAMTAOUI, A., and ABOUTAJDINE, D. Adaptive Quantization Based on Saliency Map at Frame Level of H.264/AVC Rate Control Scheme. Proceedings of European Workshop on Visual Information Processing (EUVIP). Paris: 2011. p. 61- 66. ISBN: 978-1-4577-0072-9.

[6] LEE, Sang Heon, LEE, Sang Hwa, CHO, Nam Ik. Low Bit Rates Video Coding Using Hybrid Frame Resolutions. IEEE Transactions on Consumer Electronics, [online]. December 2009, vol. 56, iss. 2 [cit. 2010-06-07]. ISSN: 0098-3063. Available at: http://

ieeexplore.ieee.org/ stamp/ stamp.jsp?tp=&arnumber=5506000.

[7] CHI, Ming-Chieh, CHEN, Mei-Juan, YEH, Chia-hung, JHU Jyong-An. Region-of-Interest Video Coding Based on Rate and Distortion Variations for H.263+. Signal Processing: Image Communication [online]. December 2007, vol. 32, iss. 2, [cit. 2008-02]. ISSN: 0923-5965. Available at: http://www.sciencedirect.com/science/article/pii/S0923596507001452#.

[8] INOUE, Masayuki, KIMATA, Hideaki, FUKAZAWA, Katsuhiko and MATSUURA Norihiko. Partial Delivery Method with Multi-Bitrates and Resolutions for Interactive Panoramic Video Streaming System. Proceedings IEEE International Conference on Consumer Electronics (ICCE). Las Vegas: 2011. p. 891- 892. ISBN: 978-1-4244-8711-0.

[9] WONG, Haohong and EL - MALEH, Khaled. Joint Adaptive Background Skipping and Weighted Bit Allocation for Wireless Video Telephony. Proceedings International Conference on Wireless Networks, Communications and Mobile Computing. Haui: 2005, p. 1243 – 1248, ISBN: 0-7803-9305-8.

[10] SHUXI, Lu, YONGCHANG, Shi and YANG, Xia. Method of Adjustable Code Based on Resolution Ratio of Spatial Domain in Surveillance Region of Interest. Proceedings International Conference on Multimedia Technology (ICMT). Ningbo: 2010. p. 1-4. ISBN: 978-1-4244-7871-2.

[11] SANCHEZ, Gustavo, SAMPAIO, Felipe and DORNELLES, Robson, AGOSTINI, Luciano. Efficiency Evaluation and Architecture Design of SSD Unities for the H.264/AVC Standard. Proceedings Programmable Logic Conference (SPL). Ipojuca: 2010. p. 171-174. ISBN 978-1-4244-6309-1.

[12] Available at: http://trace.eas.asu.edu/yuv/index.html.

[13] KAUR, Amrit, SIRCAR, Pradip and BANERJEE, Adrish. Interpolation of Lost Frames of a Video Stream Using Object Based Motion Estimation and Compensation. Proceedings IEEE India Conference (INDICON). Kanpur: 2008. p. 40-45. ISBN 978-1-4244-2747-5.

[14] PENG, Yu-Chun, CHANG, Hung-An, LIANG, Chia-Kai, CHEN, Homer and KAO, Chang-Jung. Integration of Image Stabilizer with Video Codec for Digital Video Cameras. Proceedings IEEE

International Symposium on Circuits and Systems (ISCAS'05). Kobe: 2005. p. 4871- 4874. ISBN: 0-7803-8834-8.

[15] Available at: www.ffmpeg.org

[16] MARTINEZ-RACH, Miguel, LOPEZ, Otoniel, PINOL, Pablo, PEREZ MALUMBRES, Manuel, OLIVER, Jose and CALAFATE, Carlos Tavares. Quality Assessment Metrics vs. PSNR Under Packet Loss Scenarios in MANET Wireless Networks. Proceedings International Workshop on Mobile Video. Bavaria: 2007. p. 31-36. ISBN: 978-1-59593-779-7.

[17] MARTINEZ-RACH, Miguel, LOPEZ, Otoniel, PINOL, Pablo, PEREZ MALUMBRES, Manuel, OLIVER, Jose and CALAFATE, Carlos Tavares. Behavior of Quality Assessment Metrics Under Packet Losses on Wireless Networks. XIX Jornadas de Paralelismo. Castellón: 2008.

[18] International Telecommunication Union. Methodology for the Subjective Assessment of the Quality of Television Pictures. ITU-R, Rec. BT.500-11, 2002. Available at: http://www.dii.unisi.it/~menegaz/DoctoralSchool2004/papers/ITU-R_BT.500-11.pdf.

CHAPTER EIGHT

Identifying the Position of the Motion Region of Interest to be Adapted for Video Streaming

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

Abstract

Real-time video streaming over wireless network suffers from bandwidth channel limitations that are unable to handle the high amount of video data. In this study, we propose an adaptive scheme to reduce the amount of video data by identifying and extracting the high motion region, which we call it the Region Of Interest (ROI) and drop the less motion region which is the non-Region Of Interest (non-ROI). The Sum of Absolute Differences (SAD) is computed vertically and horizontally to the consecutive video frames to identify the motion region. The videos have different characteristics and motion levels; therefor the size of the ROI will be different from one video to another. The server will identify and extract the ROI from the frames that are between reference frames; where linear interpolation is performed on the mobile side to reconstruct the ROI from the reference frames. We evaluate the videos according to the encoding size and the Mean Opinion Score (MOS) measurement. The results show that our technique significantly reduces the amount of data that are streamed over wireless network with acceptable video quality that are provided to the mobile viewers.

Keywords

Streaming Video, Sum of Absolute Differences, Reference Frames, Region of Interest , Mean Opinion Scores

8.1 Introduction

Real-time video streaming over wireless network is subject to impairments, either due to high error rate or bandwidth channel limitations. Bandwidth channel limitations are considered as the major challenge to the video stream over wireless networks [1]. Bandwidth is one of the most critical resources that should be managed efficiently in the way that could handle the amount of video traffic [2]. Therefore, it is desirable to adjust the transmission rate for the streaming video and according to the perceived congestion level to maintain the suitable losses level in wireless network.

The streaming video should be adapted according to the network bandwidth [3, 4]. Network adaptation refers to how much network resources (e.g., bandwidth) a video stream should be utilize for video content, resulting in designing an adaptive mechanism [5].

H.264/AVC allows the standard - based scalability of temporal, spatial, and quality resolution for adaptive the video frames sequence [6,7]. H.264/AVC contains a rate-control algorithm that dynamically adjusts the encoder parameters to achieve a target bit rate that allocates a budget of bits to the video frames sequence. The main concept of the rate-control algorithm is quantitative model that describe the relationship between quantization parameter (QP) and the actual bit rate [8]. QP has a great impact on the encoder performance, because it regulates on how much spatial details are saved. The QP are increases as some of the details are aggregated so that the bit rates drops with some increases in distortion and some loss of the video quality [9].

The main feature of H.264/SVC [10] is to provide bandwidth-optimized transmission for video streaming by observing current network conditions. H.264/SVC provides three types of enhancements for optimized the bandwidth transmission. First, it can support spatial enhancements of quality through a signal-noise-ratio. Second, it can support temporal enhancements by changing the frame rate, and finally it can support spatial enhancements through resolution. The H.264/SVC encode the video in the way that can be selectively

transmitted according to the type option, content and network condition by using a bit stream extractor [11].

The user attention is the ability to identify the interested parts for a given scene, called attention area or Region Of Interest (ROI) [12]. Shuxi et al. [13] adjust the resolution of the video frames by dividing the frames into ROI and non-ROI. The ROI have more details, as it is the most important region in the video frames. The ROI will perform coding on the original resolution, while the non-ROI will perform coding on low resolution. Mavlankar et al. [14] examine how to determine the slice size for streaming the ROI. The server will adapt the streaming video frames according to the regions size of the content that are desired by the client. The optimal slice size achieves the best trade-off to minimize the expected number of bits that transmitted to the client per frame and it will be depends on the display resolution of the mobile screen. Liu et al. [15] adaptive the video frames by identifying the ROI and skipping the background (non-ROI) for scalable video coding. The adaptive skipping decision depends on the motion activities of the non-ROI, while the saved bits by skipping non-ROI is used to enhance the quality of the ROI.

In this paper, we propose a video adaptation technique to reduce the amount of data to be streamed over limited bandwidth. The Sum of Absolute Differences (SAD) is computed vertically and horizontally to the consecutive video frames to identify the highest motion region within the video frame which we call it the Region Of Interest (ROI). The ROI is extracted from the frames that are between reference frames, while the non-ROI will be dropped. The size of the ROI and the non-ROI will be different from one video to another, as each video had different characteristic and different motion level. Linear interpolation is applied on the mobile side to reconstruct the non-ROI from the reference frames. Two scenarios are been considered and evaluated according to the encoding size and Mean Opinion Score (MOS) measurement. The first scenario, where the original videos are coded with a low bit rate and the second scenario is coded with a high bit rate for the proposed video streaming technique.

8.2 The Proposed Technique

To ensure a good users experience to the videos that are transmitted to the mobile device, it is necessary to adapt the streaming videos rates over limited channel bandwidth. Therefore, we propose a technique to reduce the video data by identifying and extracting the high motion region in the video frames. The motion region in the video is considered as ROI, while the less motion region is considered as non-ROI. The ROI with the reference frames are streamed to the mobile device. The mobile device will perform a reconstruction mechanism to the non-ROI pixels that are dropped on the server side, and according to the following steps:

8.2.1 Identifying the ROI

The streaming server will establish the connection according to the mobile request. The server will identify the position of the highest motion region (ROI) in the frames and drop the less motion region (non-ROI). The technique is used to identify the ROI is the Sum of Absolute Differences (SAD). The SAD is a commonly used technique for motion estimation in various video coding standards like H.264 [16]. The SAD value will be low except for the changes induced by objects moving between frames. If there is a lot of motion in the video frames, the SAD value will be relatively high, and if there is less motion then the SAD value will be less.

The consecutive video frames are scanned vertically and horizontally by computing the SAD in two ways to identify the highest intra-column differences and the highest intra-row differences within the video frames:

- Vertically: The SADv(x) is computed column-by-column to the consecutive video frames and from left-to-right to identify the highest width value, as shown in Figure 1 and according to (1).

$$\mathrm{SAD_V}\ (x) = \sum_{j=0}^{N-1} \sum_{z=1}^{L-1} |\mathrm{F_z}(x, \mathrm{j}) - \mathrm{F_{z-1}}(x, \mathrm{j})| \qquad (1)$$

- Horizontally: The $\mathrm{SAD_H(y)}$ is computed row-by-row to the consecutive video frames and from top-to-bottom to identify the highest height value, as shown in Figure 2 and according to (2).

$$\mathrm{SAD_H}\ (y) = \sum_{i=0}^{M-1} \sum_{z=1}^{L-1} |\mathrm{F_z}(i, \mathrm{y}) - \mathrm{F_{z-1}}(i, \mathrm{y})| \qquad (2)$$

To identify the size and the position of the ROI, the average value to the results that is obtained from (1) is calculated according to (3), while the average value to the results that is obtained from (2) is calculated according to (4).

$$\mathrm{Avg}(x) = \frac{1}{M} * \sum_{i=0}^{M-1} \mathrm{SAD_V}\ (x_i) \qquad (3)$$

$$\mathrm{Avg}(y) = \frac{1}{N} * \sum_{j=0}^{N-1} \mathrm{SAD_H}\ (y_j) \qquad (4)$$

where L is the length of the frame sequences, N is the height and M is the width.

The average values that are obtained from equations 3 and 4 will be crossed with the intra-column and the intra-row differences. The crossing point represents the coordination points within the video frames to identify the size and the position of the highest motion.

Figure 1: Scanning the consecutive video frames based on SADV(*x*).

Figure 2: Scanning the consecutive video frames based on SADH(*y*).

The test videos are used in this work were the samples of video sequences, with a resolution of 176 *x* 144 pixels [17]. The chosen videos are well known as professional test videos that have different characteristics and different motion levels, as shown in Table I.

The SADV and the SADH are calculated to the consecutive video frames to find the intra-columns differences and the intra-rows differences values to the test videos. The intra-differences are used to identify the most differences within the video frames. The crossing points that are obtained from calculating the average values for SADV and SADH are used to identify the position and the size of the motion region. The motion level is varying within the video frames, and the motion activities will be different from one video to another as each video had characteristic, as shown in Figures 3-14.

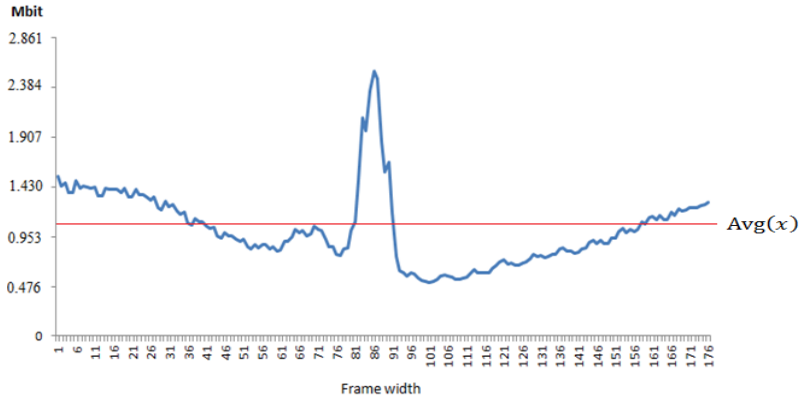
For Waterfall and Foreman videos, the values that are obtained from SADV are the lowest in the frames sides, while it is the highest in the middle of the frames, as shown in Figures 3(a) and 5(a), respectively. The values that are obtained from SADH are very hard to estimate for both videos; therefore it is considered the complete frame height, as shown in Figures 3(b) and 5(b), respectively. The reason for that is, the Waterfall video is zooming out all the time, while the Foreman video is shaking all the time. Since each video had different characteristic, therefore the size of the ROI for both videos are different, as shown in Figures 14 (a) and (c), respectively.

Table 1: The description of the test videos and their ROI size.

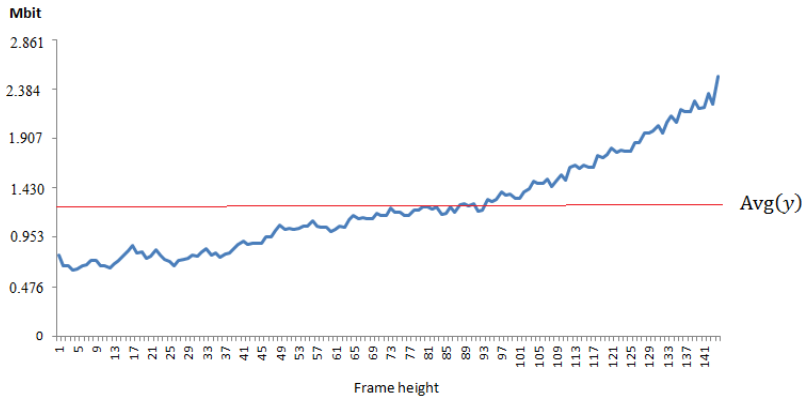
Test Videos	Number of Frames	Motion Level	ROI%	Description
Waterfall	300	Medium	9,66	A water is falling down with zooming out video
News	300	Low	28,61	A man and a woman are reporting the news with background dance
Foreman	300	High	43,75	A constructor builder is talking with a weaving hand and the video is shaking
Akiyo	300	Low	13,89	A woman is reporting the news
Silent	359	Medium	21,29	A woman is demonstrating a sign language
Coastguard	359	Medium	40,28	A coastguard boat moving in the river
Container	359	Low	12,50	A ship are carrying containers and moving slowly in the sea
Mobile	359	High	49,31	A train toy is pushing a ball with a moving calendar
Carphone	458	High	55,56	A man is talking inside a moving car
Claire	592	Low	12,05	A woman is reporting the news
Highway	2000	High	40,97	A car is driven in the highway

For News, Akiyo, Silent, Container, and Claire videos. The values that are obtained from SAD_H and SAD_V are mostly the highest in the middle of the frames, as shown in Figures 4, 6, 7, 9 and 12, respectively. Since the motion activates for these videos are different, therefore the position and the size of the ROI are different as well, as shown in Figures 14 (b), (d), (e), (g), and (j), respectively.

For Coastguard, Mobile, Carphone, and Highway videos. The SAD_H values are the highest in the middle of the frames, as shown in Figures 8(b), 10(b), 11(b), and 13(b), respectively. The SAD_V value is very hard to estimate and the reason for that is, the videos had a dynamic background, therefore it is considered the complete frame width as shown in Figures 8(a), 10(a), 11(a), and 13(a), respectively. The ROI size and position for these videos are different from one video to another as the motion level are different and in different position, as shown in Figures 14 (f), (h), (i), and (k), respectively.

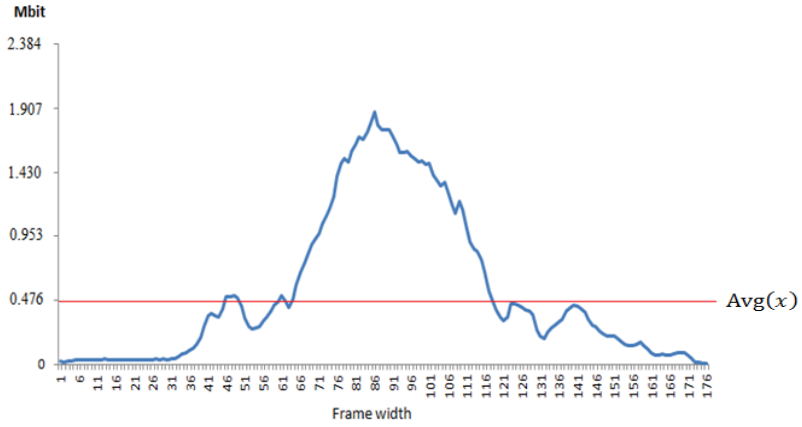


a. The SADV (*x*)

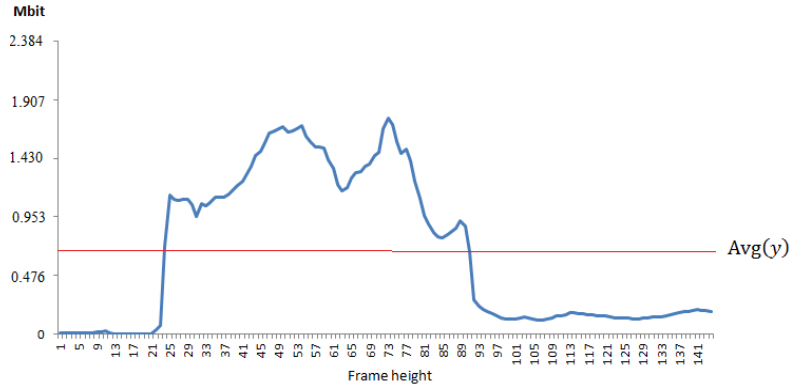


b. The SADH (*y*)

Figure 3: Waterfall video.

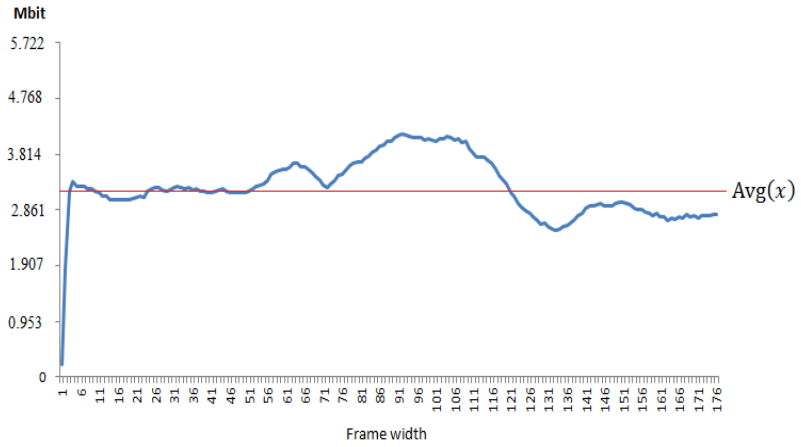


a. The SADV (*x*)

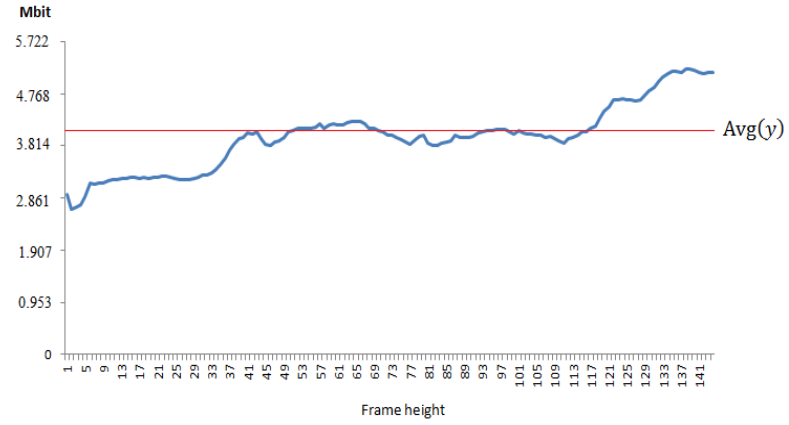


b. The SADH (*y*)

Figure 4: News video.

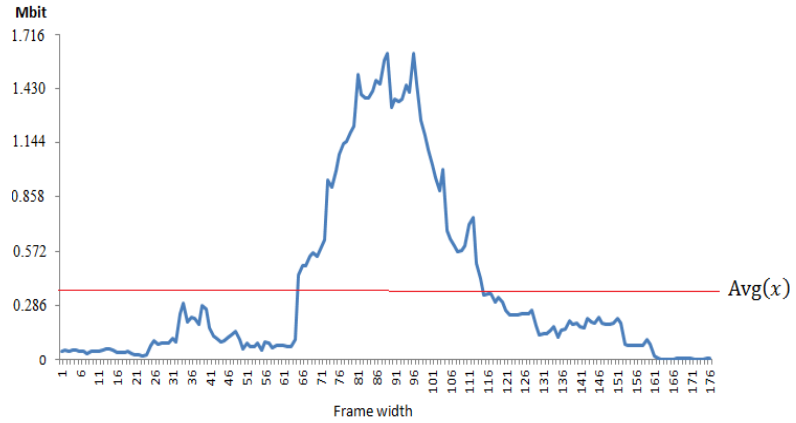


a. The SADV (x)

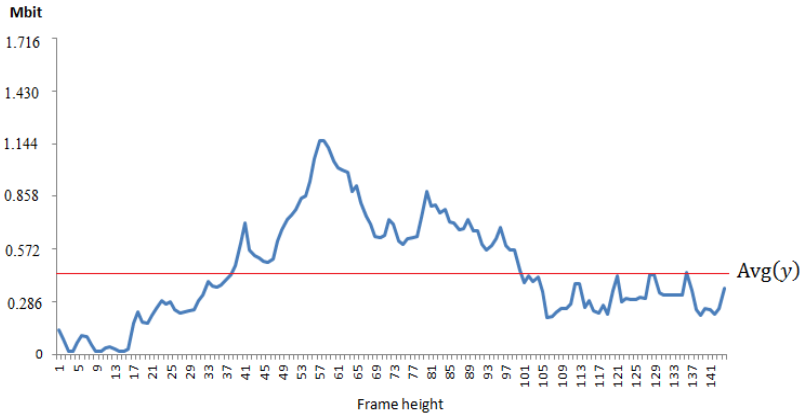


b. The SADH (y)

Figure 5: Foreman video.

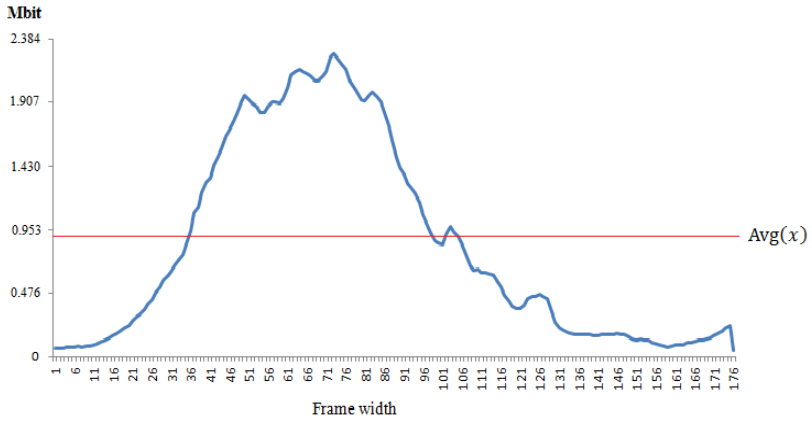


a. The SADV (*x*)

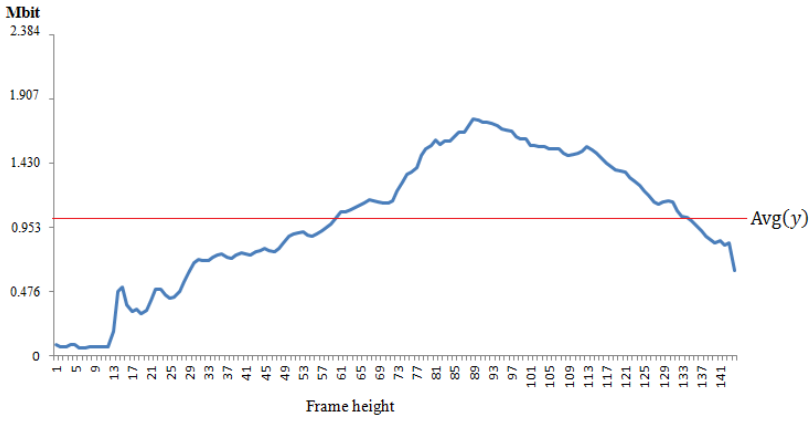


b. The SADH (*y*)

Figure 6: Akiyo video.

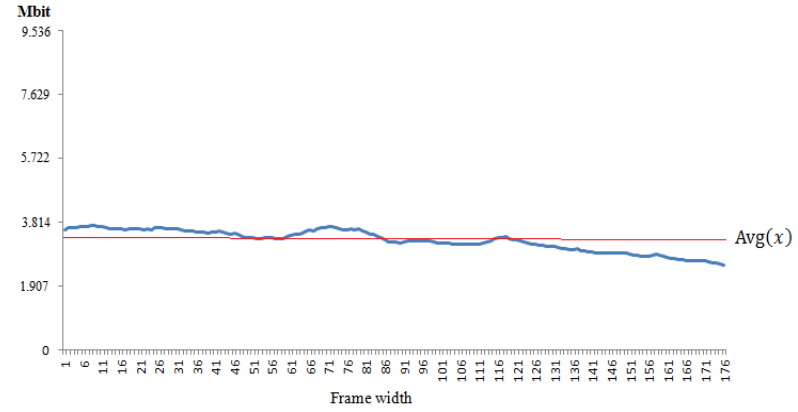


a. The SADV (*x*)

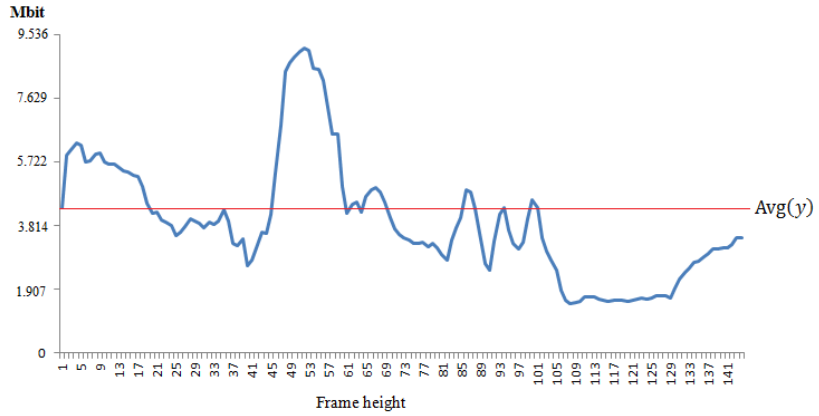


b. The SADH (*y*)

Figure 7: Silent video.

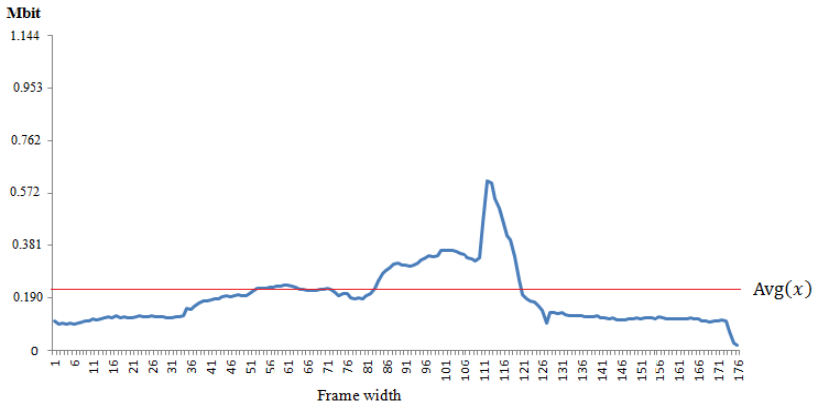


a. The SADV (*x*)

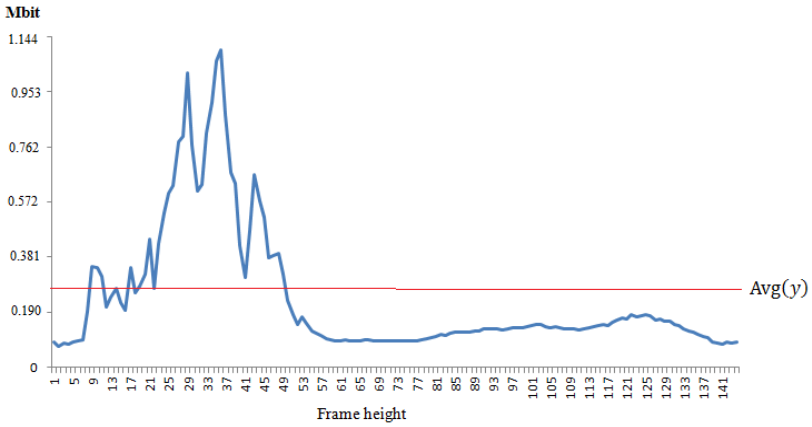


b. The SADH (*y*)

Figure 8: Coastguard video.

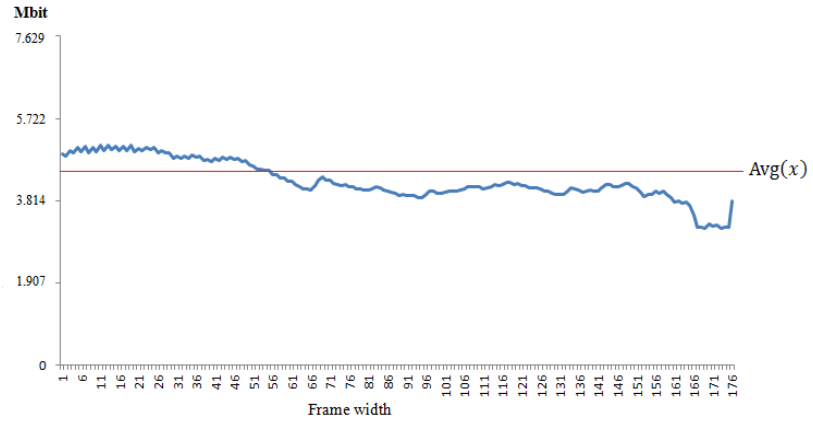


a. The SADV (*x*)

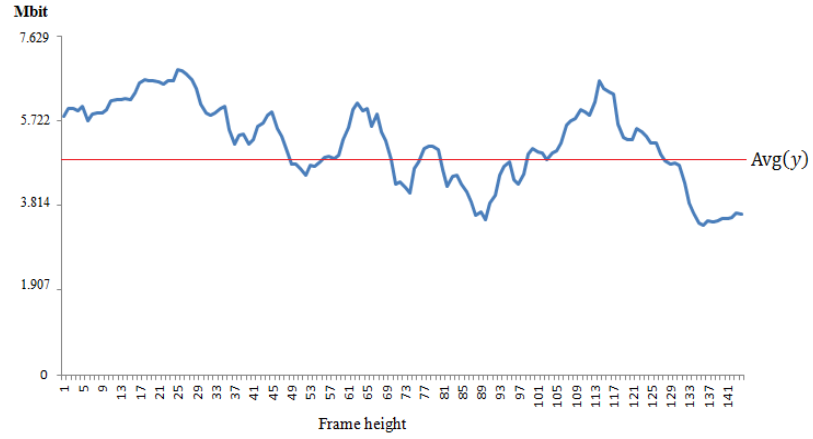


b. The SADH (*y*)

Figure 9: Container video.

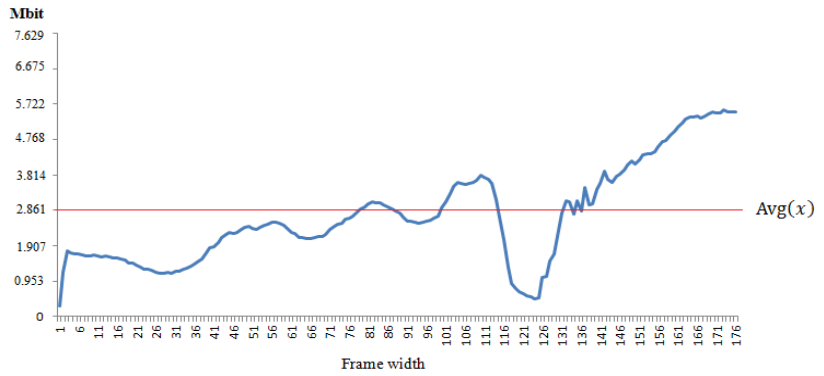


a. The SADV (*x*)

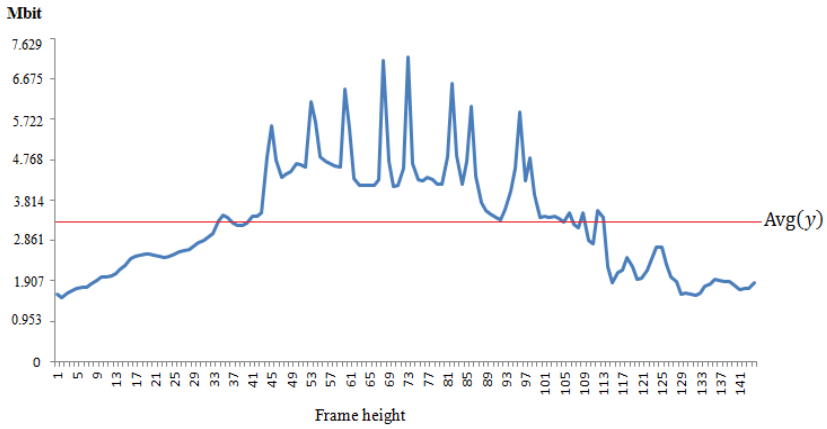


b. The SADH (*y*)

Figure 10: Mobile video.

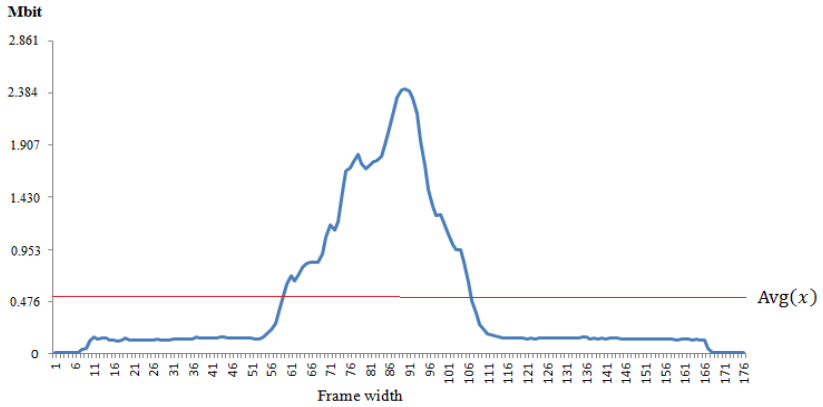


a. The SADV (*x*)

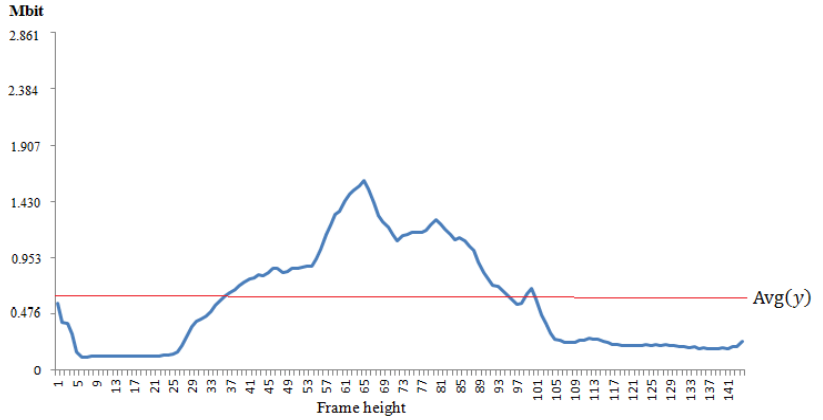


b. The SADH (*y*)

Figure 11: Carphone video.

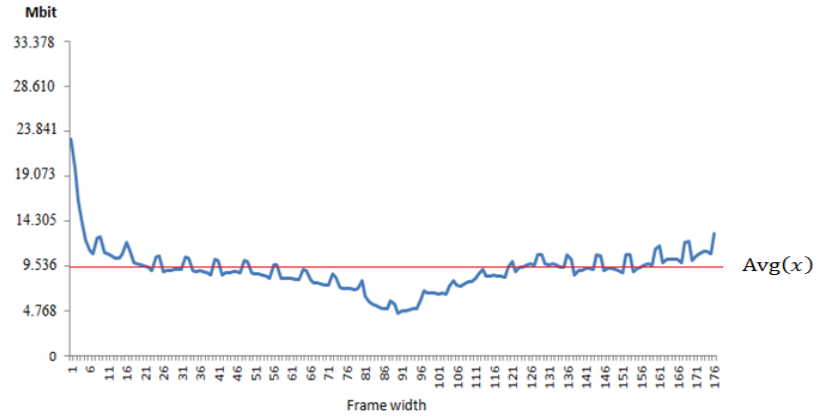


a. The SADV (x)

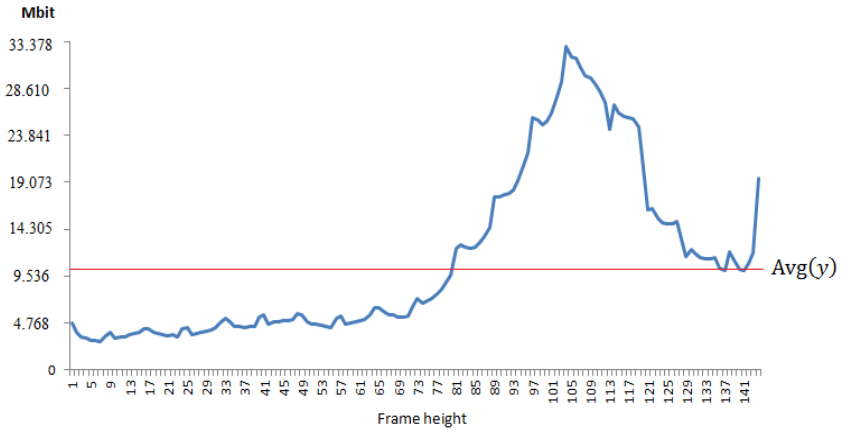


b. The SADH (y)

Figure 12: Claire video.

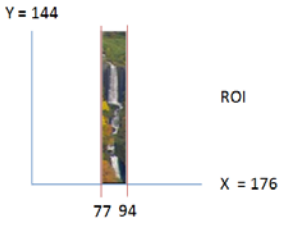


a. The SADV (x)

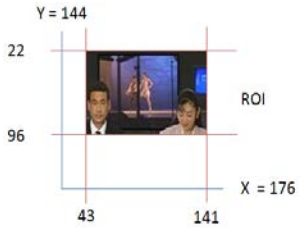


b. The SADH (y)

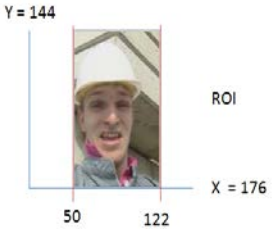
Figure 13: Highway video.



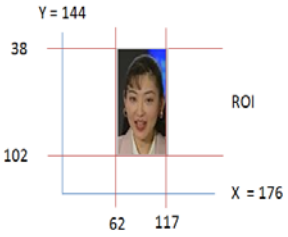
a. Waterfall



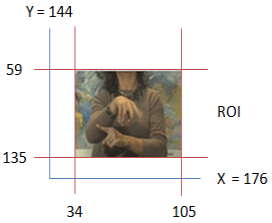
b. News



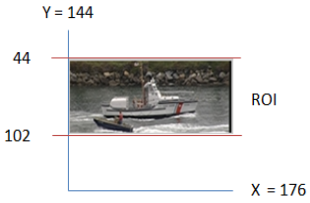
c. Foreman



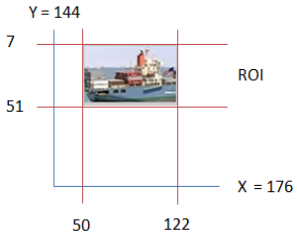
d. Akiyo



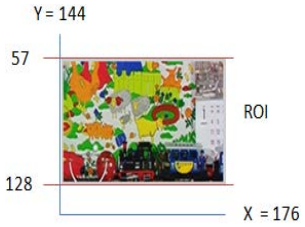
e. Silent



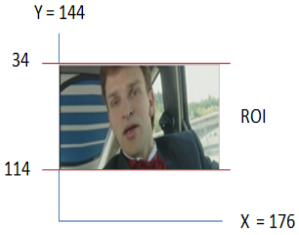
f. Coastguard



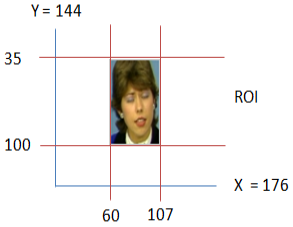
g. Container



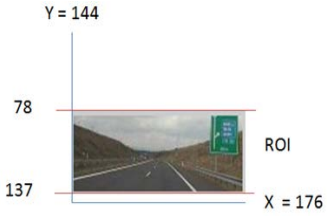
h. Mobile



i. Carphone



j. Claire



k. Highway

Figure 14: The ROI for the test videos.

The values that are obtained from SAD_V and SAD_H, and their averages are used to identify the highest changes in the consecutive video frames. The highest changes that are between the coordination points is considered as the ROI, where the ROI size and position are different from one video to another, as shown in Table I and Figure 14.

8.2.2 Extracting the ROI

The server will compute the SAD vertically and horizontally and their averages to the consecutive video frames to determine the position and the size of the ROI. The video will be splitted into two parts. The first part is the reference frames, where the reference frames are the complete frame. The reference frames are set in this study is every fifth frame, as the maximum distance between frames that do not have high effect on the quality of the viewers' perception [18]. The second part is the ROI that are extracted from the frames that are between reference frames and drop the non-ROI pixels, as shown in Figure 14. The reference frames with the ROIs are encoded by using H.264 [19] and it will be transmitted to the mobile device.

8.2.3 Reconstructing the Video Frames

The mobile devices start receiving the video frames (reference frames and the ROIs). Linear interpolation [20] is performed to replace the surrounding pixels to the ROI that been dropped in the server side. The dropped pixels that are related to the ROI will be reconstructed from the reference frames to return the frames to their original resolution. The mobile device will start playing the video according to the playout rate.

8.3 Quantization Parameter Adaptation

The video is encoded to obtain the optimum visual quality within the available bandwidth. The bit allocation for the video should achieve the trade-off between encoding the video quality and bandwidth limitations. The bit allocation for the video is encoded by H.264 ffmpeg

codec [19]. The videos are encoded to identify the effectiveness of the bit rates and the quantization parameter (QP) on the encoding size.

Two scenarios are proposed to encode the videos, in the first scenario; where the original videos are encoded with a bit rate of 64 kbps. In the second scenario (the proposed scenario), where the reference frames and the ROIs are encoded with a bit rate of 128 kbps. The videos for the first scenario are encoded with a default QP, while the QP in the second scenario has been set to achieve equivalents encoding size to the videos that are in the first scenario. The idea for encoding the videos in the second scenario with a high bit rate and adaptive QP parameter is not only to increase the quality of the videos, it is also to get equivalents encoding size to the videos that are in the first scenario, that can be transmitted over low bandwidth.

Table 2: Video encoding for the two scenarios.

Test Videos	Size (KB), QP=2	Bitrate: 64	Size (KB)	Coding Efficiency Gain	Bitrate: 128	Size (KB)	Coding Efficiency Gain
Waterfall	1280	10	165	**87,10**	3	138	**89,21**
News	1128	11	167	**85,19**	6	179	**84,13**
Foreman	616	12	146	**76,29**	11	144	**76,62**
Akiyo	350	6	165	**52,85**	3	169	**51,71**
Silent	664	7	200	**69,88**	4	216	**67,47**
Coastguard	1390	13	177	**87,27**	10	177	**87,27**
Container	547	6	219	**59,96**	3	205	**62,52**
Mobile	3275	24	207	**93,68**	17	198	**93,95**
Carphone	1370	13	221	**83,87**	10	218	**84,09**
Claire	551	3	301	**45,37**	3	271	**50,82**
Highway	4564	14	707	**84,50**	10	640	**85,97**

The efficiency coding gain percentage that are calculated from encoding the original videos, and the videos for the first scenario with a default QP and the videos for the second scenario with adaptive QP is to get equivalent encoding size. The encoding sizes are different from one video to another, as the videos had different characteristics, as shown in Table II.

8.4 Subjective Viewing Test

8.4.1 Test Methods

It is well known that the peak Signal-to-Noise-Ratio (PSNR) does not always rank the 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 [21]. One of the most reliable ways of assessing the quality of a video is subjective evaluation of 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 [22].

8.4.2 Testing Materials and Environments

The videos are displayed on a 17 inch FlexScan S2201W LCD computer display monitor of type EIZO with a native resolution of 1680 x 1050 pixels. The videos are displayed with resolution of 176 x 144 pixels in the centre of the screen with a black background.

The MOS measurement is used in this study to evaluate the video quality according to the guidelines outlined in the BT.500-11 recommendation of the radio communication sector of the International Telecommunication Union (ITU-R). We use a lab with controlled lighting and set-up according to the ITU-R recommendation. The score grades in this methods range from 0 to 100. These ratings are mapped to a 5-grade discrete category scale labelled with Excellent, Good, Fair, Poor and Bad [23].

The subjective experiment was conducted at Blekinge Institute of Technology in Sweden. The participant of thirty non-expert test subjects, 26 males and 4 females, they were all university students and their ages range from 20 to 40.

The users observed two scenarios for displaying the videos; in the first scenario, where the observer evaluates the original video with a low bit rate and for the second scenario (the proposed scenario) the observer evaluates the reconstructed videos with a high bit rate.

The amounts of data are gathered from the subjective experiments with respect to the opinion scores that were given by the individual viewers. Concise representation of the data is achieved by calculating the conventional statistics such as the mean score and 95% confidence interval [23].

8.5 Experiment Results

A panel of users evaluates the two scenarios according to Mean Opinion Score (MOS) measurement, as shown in Figure 15. In the first scenario, the original videos are decoded with a bit rate of 64 kbps. The second scenario (the proposed scenario), the videos are decoded with a bit rate of 128 kbps and the observers evaluate the videos after the non-ROI is been reconstructed.

For Waterfall videos, the average score for both scenarios are more than three. The score for the second scenario, it is show better score than the first scenario, as the observers didn't notice the interpolation effected on the video, as the background pixels (non-ROI) with the ROI had similar data, although the Waterfall video is zooming out all the time.

For Coastguard, Mobile, Carphone, and Highway videos, the user's average score is less than three for both scenarios. The first scenario show better score than the second scenario, as the observers notice the effect of interpolation on the dynamic background (non-ROI), although the second scenario are decoded with a high bit rate.

For News, Akiyo, Silent, Container, and Claire videos, the user's average score is more than three for both scenarios, as an indicator that the interpolation has less effect on the videos. The non-ROI (background) in these videos are static except the Container video, as the background pixels is quite similar to the ROI with little changes among the frames pixels, therefore the observers did not notice the interpolation effect on such video.

For Foreman videos, the average score for both scenarios are less than three. The original video in the first scenario that are decoded with a low bit rate show better score than our proposed scenario, although the second scenario is decoded with a high bit rate. The reason for that is, Foreman video is shaking all the time with scene changes, therefore the video been effected by interpolation and for this reason the observers gives the lowest score.

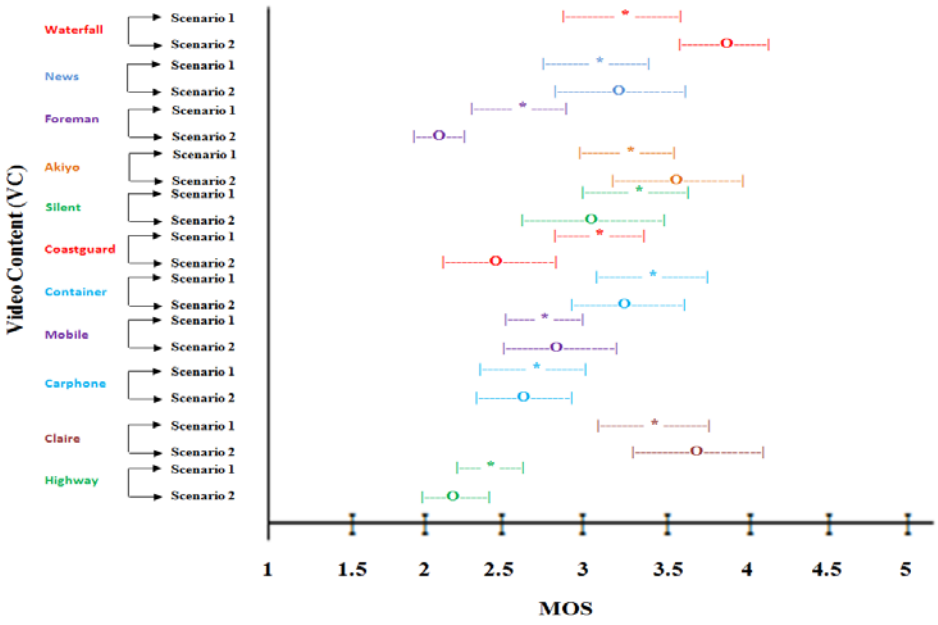


Figure 15: The MOS for different videos content and for two scenarios.

It is been notice from Table I and Figure 15. The proposed scenario (second scenario), if the ROI size is more than 40% from the original frame, then the score of the MOS is less than three, and If the ROI size is less than 40% then the MOS is more than three.

It can be summarised from that, the smallest ROI size that are extracted from the video frames, it is the highest score that we could have from the user's panel evaluation.

8.6 Conclusion

In this study, we proposed an adaptive scheme to reduce the amount of video data that are streamed to the mobile device over wireless network. The proposed adaption scheme is to identify, extract, the ROI from the frames that are between reference frames. The reference frames is set as every fifth frame in the video sequence. The Sum of Absolute Differences (SAD) is used to identify the highest motion region in the consecutive video frames. The highest motion region is considered as the Region Of Interest (ROI). The receiving video stream by the mobile device, will reconstruct the non-ROI by using linear interpolation and from the reference frames.

It is been notice from that, the MOS is high for the videos with statics background, with exception to the dynamic videos that had a similar background with the ROI, e.g., Container and Waterfall videos. The MOS for the dynamic background its show acceptable rate, while the videos that are shaking, e.g., Foreman video shows the worst evaluation from the user's panel, as the reconstruction mechanism had high effect on the video.

From the results, it is shows that the user's opinions are linked to the SAD, where the SAD is used to identify the position of the motion region (ROI). The user's panel evaluate the videos with high scores for the statics background that has a small ROI size.

Adapting the video stream by extracting the ROI and skipping the non-ROI was proposed jointly with a high bit rate coding to achieve better quality to the motion region and to reduce the size of the video that are transmitted over limited bandwidth. The background of the

videos quality is degraded and especially for the videos with dynamic background; but it could be a satisfactory approach that can be provided to the mobile users.

References

[1] D. Tian, X. Li, G. Al-Regib, Y. Altunbasak, and J. R. Jackson,“ Optimal packet scheduling for wireless video streaming with error-prone feedback,” In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC ´04), pp. 1287-1292, March, 2004.

[2] 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, pp. 1566 – 1574, 2003.

[3] 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,” the 6th IEEE International Conference on Computer and Information Technology (CIT’06), 2006.

[4] H. Luo, M-L. Shyu, and S-C. Chen,“An end-to-end video transmission framework with efficient bandwidth utilization,” the IEEE International Conference on Multimedia and Expo. (ICME’04), pp. 623-626, June 2004.

[5] F. Yang, Q. Zhang, W. Zhu, and Y-Q. Zhang,“ Bit allocation for scalable video streaming over mobile wireless internet,” the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 2142-2151, March 2004.

[6] L. Ferreira, L. Cruz, and P. A. Amado Assunção, “ H.264/SVC ROI encoding with spatial scalability,” the International Conference on Signal Processing and Multimedia Applications, Porto, Portugal, July 26-29, pp. 212-215, 2008.

[7] T. Schierl, T. Stockhammer, and T. Wiegand,“Mobile Video Transmission Using scalable video coding,” IEEE Transactions on Circuits and System for video Technology, Vol. 17, No.9, September, 2007, pp. 1204 – 1217.

[8] L. Czuni, G. Csaszar, and A. Licsar, "Estimating the optimal quantization parameter in H.264," 18th International Conference on Pattern Recognition (ICPR'06), pp. 330-333, August 2006.

[9] M. Hrarti, H. Saadane, M. Larabi, A. Tamtaoui, D. Aboutajdine," Adaptive quantization based on saliency map at frame level of H.264/AVC rate control scheme," 3rd European Workshop on Visual Information Processing (EUVIP '11), pp. 61- 66, July 2011.

[10] J.-H. Lee, and C. Yoo, "Scalable ROI algorithm for H.264/SVC-based video streaming," In Proceedings of the IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, pp. 882-887, May 2011.

[11] J-H. Lee, and C. Yoo,"Scalable ROI algorithm for H.264/SVC-based video streaming," the IEEE Transactions on Consumer Electronics, vol. 57, pp. 882-887, 2011.

[12] M. Inoue, H. Kimata, K. Fukazawa, and N. Matsuura, "Partial delivery method with multi-bitrates and resolutions for interactive panoramic video streaming system," IEEE International Conference on Consumer Electronics (ICCE'11), pp. 891- 892, January 2011.

[13] Shuxi, L., Yongchang, S. Yang, X., "Method of Adjustable Code Based on Resolution Ratio of Spatial Domain in Surveillance Region of Interest. In: Int. Conf. on Multimedia Technology (ICMT), pp. 1-4 (2010).

[14] A. Mavlankar, P. Baccichet, D. Varodayan, and B. Girod, " Optimal slice size for streaming regions of high resolution video with virtual Pan/Tilt/Zoom functionality," the 15th European Signal Processing Conference (EUSIPCO'07), September 2007.

[15] Q. Liu, R-M. Hu, and Z. Han,"Adaptive background skipping algorithm for region-of-interest scalable video coding," 11th IEEE Singapore International Conference on Communication Systems (ICCS '08), pp. 788 – 792, Novembers 2008.

[16] A. Dimou, O. Nemethova, and M. Rupp, " Scene change detection for H.264 using dynamic threshold techniques," the 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Service, July 2005.

[17] trace.eas.asu.edu/yuv/index.html.

[18] A. Kaur, P. Sircar, A. Banerjee,"Interpolation of lost frames of a video stream using object based motion estimation and compensation," IEEE India Conference (INDICON'08), pp. 40-45, December 2008.

[19] ww w.ffmpeg.org

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

[21] 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," Proceedings of the International Workshop on Mobile Video, pp. 31-36, June. 2007.

[22] M. Martinez-Rach, O. López, P. Piñol, M.P. Malumbres, J. Oliver, and C.T. Calafate,"Behavior of quality assessment metrics under packet losses on wireless networks," XIX Jornadas de Paralelismo, Castellón, September 2008.

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

ABSTRACT

The main objective of this thesis is to provide a smooth video playout on the mobile device over wireless networks. The parameters that specify the wireless channel include: bandwidth variation, frame losses, and outage time. These parameters may affect the quality of the video negatively, and the mobile users may notice sudden stops during the playout video, i.e., the picture is momentarily frozen, followed by a jump from one scene to a different one.

This thesis focuses on eliminating frozen pictures and reducing the amount of video data that need to be transmitted. In order to eliminate frozen scenes on the mobile screen, we propose three different techniques. In the first technique, the video frames are split into sub-frames; these sub-frames are streamed over different channels. In the second technique the sub-frames will be "crossed" and sent together with other sub-frames that are from different positions in the streaming video sequence. If some sub-frames are lost during the transmission a reconstruction mechanism will be applied on 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 losing a sequence of neighbouring frames. A missing frame from the streaming video will be reconstructed based on the surrounding frames on the mobile device.

In order to reduce the amount of video data that are streamed over limited bandwidth channels, we propose two different techniques. These two techniques are based on identifying and extracting a high motion region of the video frames. We call this the Region Of Interest (ROI); the other parts of the video frames are called the non-Region Of Interest (non-ROI). The ROI is transmitted with high quality, whereas the non-ROI is interpolated from a number of references frames. In the first technique the ROI is a fixed size region; we considered four different types of ROI and three different scenarios. The scenarios are based on the position of the reference frames in the streaming frame sequence. In the second technique the ROI is identified based on the motion in the video frames, therefore the size, position, and shape of the ROI will be different from one video to another according to the video characteristic. The videos are coded using ffmpeg to study the effect of the proposed techniques on the encoding size.

Subjective and objective metrics are used to measure the quality level of the reconstructed videos that are obtained from the proposed techniques. Mean Opinion Score (MOS) measurements are used as a subjective metric based on human opinions, while for objective metric the Structural Similarity (SSIM) index is used to compare the similarity between the original frames and the reconstructed frames.

ISSN 1653-2090

2013:08

ISBN: 978-91-7295-258-4