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

Hussein Muzahim Aziz, Markus Fiedler, Håkan Grahn, and Lars Lundberg
School of Computing, Blekinge Institute of Technology, Sweden.
{hussein.aziz, markus.fiedler, hakan.grahn, lars.lundberg}@bth.se

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

I. INTRODUCTION

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

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

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

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

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

II. BACKGROUND AND RELATED WORK

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

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

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

Zheng et al. [12] proposed a scheme that integrates MDC, hybrid space-time coding structure for robust video transmission over MIMO-OFDM system. The MDC will generate multiple bitstreams of equal importance which are very suitable for multiple-antennas system. They considered a MIMO system with 4 transmitters and 4 receiver antennas for robust video transmission, thus transmitting signals for different subcarriers simultaneously over all transmit antennas. Data partition is used to divide the encoder signal bitstream into two components. In this scheme, each description is further divided into two partitions: motion vectors and Group-of-Pictures (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 a 4x4 MIMO architecture.

III. THE PROPOSED TECHNIQUE

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

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

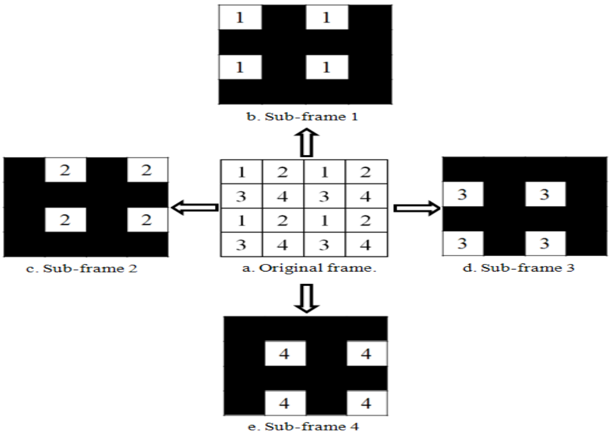


Figure 1. Frame splitting

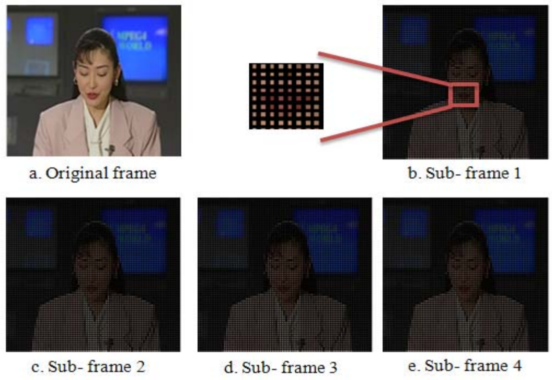


Figure 2. Snapshot of Akiyo frame splitting

A. *Encoding the sub-frames*

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

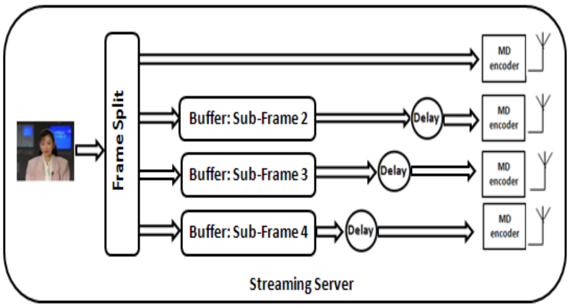


Figure 3. Streaming the sub-frames over multichannels

B. *Decoding the sub-frames*

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

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

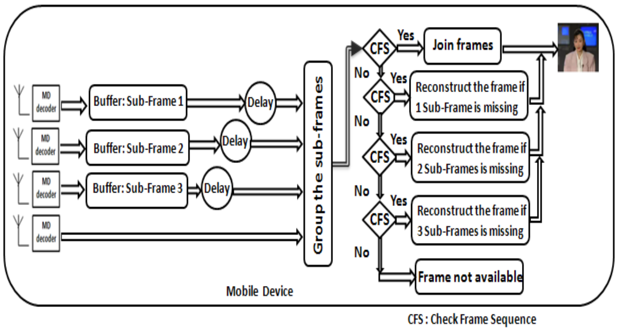


Figure 4. Receiving the sub-frames of the video in the mobile device

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

- The first CFS will check whether all the sub-frames that are related to the same original frame are available. If the four related sub-frames are available then a joining mechanism will be applied to return the frame to its original shape.
- The second CFS will check if there are at least 3 sub-frames are available as shown in Figure 5. If one sub-frame is missing then the average of the neighbouring pixels will be calculated to replace the missing frame pixels and return the full frame to its normal shape.
- The third CFS will check if there are at least 2 sub-frames available as shown in Figure 6. If two sub-frames are missing then the average of the neighbouring pixel will be calculated to replace the missing sub-frame pixels and return the full frame to its normal shape.
- The fourth CFS will check if there is at least 1 sub-frame is available as shown in Figure 1. If three sub-frames are missing then the average of the neighbouring pixel will be calculated twice, the first time to find the half of the frame as shown in Figure 6 and the second time to return the full frame to its normal shape.

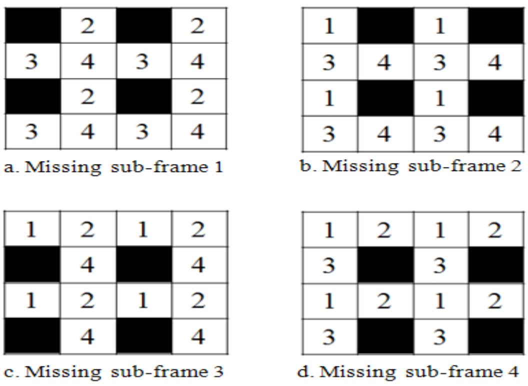


Figure 5. The possibility of missing one sub-frame

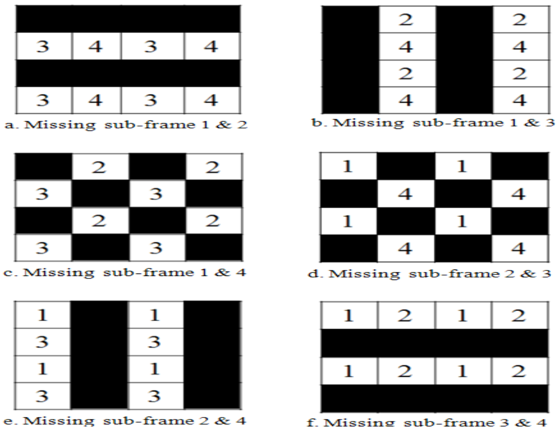


Figure 6. The possibility of missing two sub-frames

IV. SUBJECTIVE VIEWING TEST

A. Testing Methods

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

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

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

B. 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 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.

V. EXPERIMENTAL RESULTS

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

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

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

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

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 the 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.

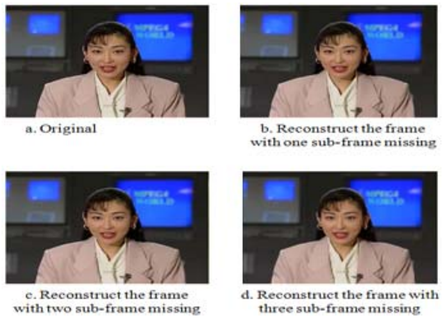


Figure 7. Snapshot of Akiyo video frame number 140

Figure 8. Snapshot of Foreman video frame number 140

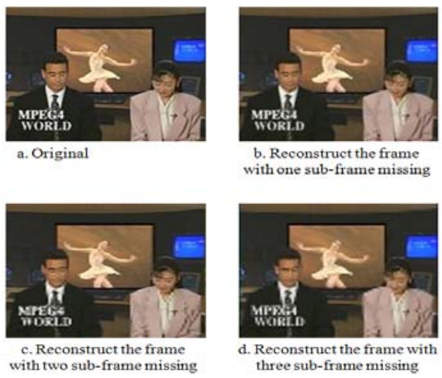


Figure 9. Snapshot of News video frame number 140

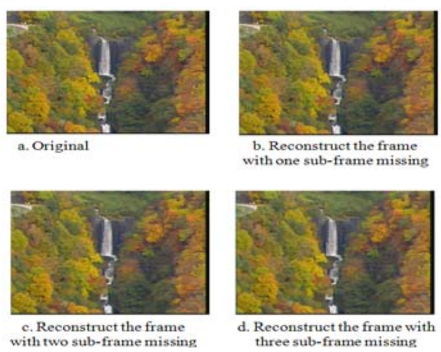


Figure 10. Snapshot of Waterfall video frame number 140

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

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

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

VI. CONCLUSIONS

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

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

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

ACKNOWLEDGMENTS

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

REFERENCES

[1] S. Kopf, T. King, F. Lampi, and W. Effelsberg, “Video color adaptation for mobile devices,” In Proc. of the 14th ACM International Conference on Multimedia, pp. 963-964, Oct. 2006.

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

[3] C. Xiaozhen, B. Guangwei, and W. Carey, “Media streaming performance in a portable wireless classroom network,” In Proc. of IASTED European Workshop on Internet Multimedia Systems and Applications, EuroIMSA ´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 Proceedings of Wireless Networking Symposium,WNCG` 03. Austin, Texas, October 26-28, 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. on Wireless Personal Multimedia Communications (WPMC’06), pp. 1-5, Sep. 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 International Multi-Symposium of Computer and Computational Sciences, IMSCCS ’06. Hanzhou, Zhejiang, IMSCCS ’06, vol. 2, pp. 257-261, June 20-24, 2006.

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

[9] W. Dapeng, Y. T. Hou, W. Zhu, Y-Q Zhang, and J. M. Peha, “Streaming video over the internet: approaches and directions,” IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Streaming Video, vol. 11, no. 3, pp. 282-300, 2001.

[10] Y. Wang, A. R. Reibman, S. Lin, “Multiple description coding for video delivery,” In Proceedings of the IEEE Journal , vol. 93, no. 1, pp. 57-70, 2005.

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

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

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

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

[15] J. G. Apostolopoulos, “Reliable video communication over lose packet networks using multiple state encoding and path diversity,” In Proceedings of the Visual Communications and Image Processing, San Jose, CA, pp. 392-409, January 24-26, 2001.

[16] H. M. Aziz, H. Grahn, and L. Lundberg, “Eliminating the freezing frames for the mobile user over unreliable wireless networks,” In Proc. of the 6th Int’l Conf. on Mobile Technology, Applications and Systems (ACM Mobility), pp. 57:1--57:4, Sep. 2009.

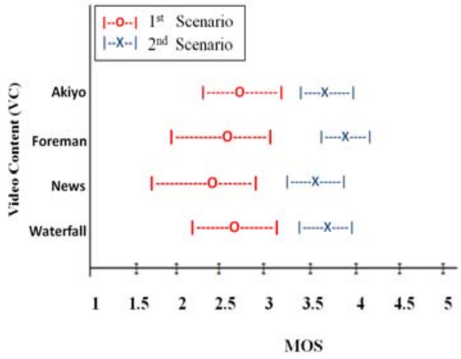
[17] http://www.ffmpeg.org.

[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, Sep. 2007.

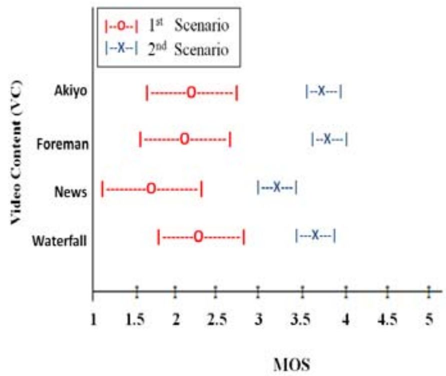
[19] M.Martinez-Rach, O.Lopez, P.Pinol, M.P. Malumbres, J. Oliver, Carlos T. Calafate, “Behavior of quality assessment metrics under packet losses on Wireless Networks,” XIX Jornadas de Paralelismo, Castellón, Spain, September 17-19, 2008.

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

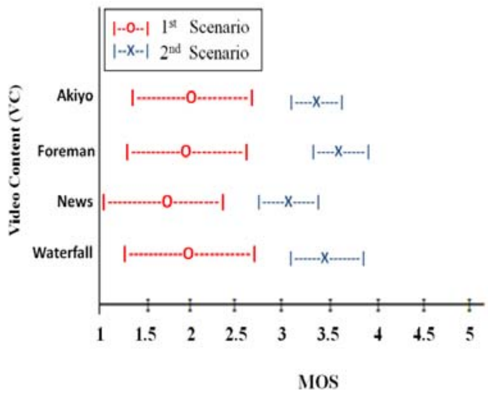
[21] http://trace.eas.asu.edu/yuv/index.html.



a. Light drop rate

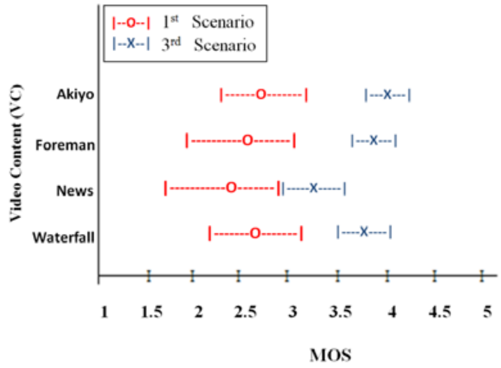


b. Medium drop rate

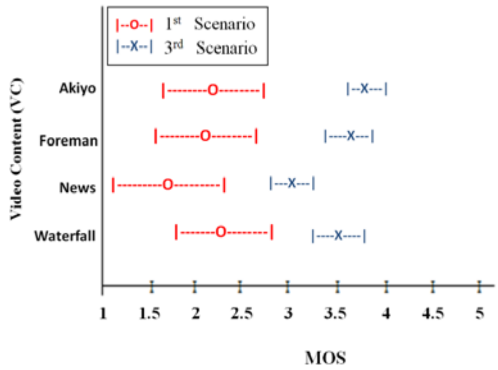


c. High drop rate

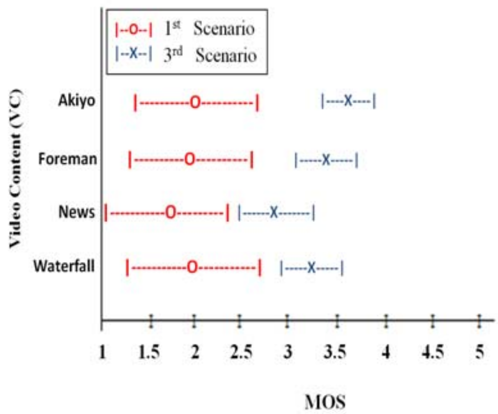
Figure 11. The MOS comparison between scenario 1 and 2 and for different video contents under different frame dropping rates. The MOS scale is from 1 (Bad) to 5 (Excellent).



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. The MOS scale is from 1 (Bad) to 5 (Excellent).