ANALYSIS OF THE EFFECTS OF PACKET LOSS AND DELAY JITTER ON MPEG-4 VIDEO QUALITY

Muhammad Saleem Koul

Dept. of Electrical Engineering, The University of Texas at Arlington, Arlington, Texas 76019
Email: mskoul@uta.edu

ABSTRACT

The MPEG-4 standard brings the promise of providing real time and streaming multimedia over broadband wireless networks. The current developments in the Third Generation Broadband CDMA technology, namely CDMA-2000 and EV-DO [1] provide us with a backbone for such real time multimedia applications. There is however a caveat associated with video transmission over RF, i.e. the existence of packet loss and jitter (Packet delay-variance). This paper analyzes recent trends in evaluating the effects of real and simulated RF packet delay and loss scenarios on MPEG-4 transmissions. Several novel approaches to Objective Video Quality Estimation are used on these results. Comparative conclusions are drawn based on the amount of Packet loss or jitter observed at the particular MPEG-4 frames.

1. INTRODUCTION

This review paper focuses on current trends in developing frameworks for Video Quality Estimation suited for lossy wireless broadband networks. We analyze a strategy represented in [2], that uses a Framework originally designed by [3]. Then this strategy is modified to suit our application in CDMA 2000 and EvDO. We also modify the strategy in the area of Video Quality Assessment by doing a comparative analysis of several major Video Quality assessment methodologies. The paper also falls short in terms of future applications. We introduce a practical application of using this methodology as a full reference Video Quality assessment suite that can be utilized by Service Providers or Wireless network vendors to assess the quality of service (QoS) measures of their multimedia transmissions. The paper goes on to introduce the UMTS (Universal Mobile Telecommunication System). UMTS is one of the first systems to offer wireless broadband over the Internet Protocol. Data rates offered are around 384 kbps in a highly mobile Wide Area Network. It also discusses the emergence of Video Streaming as one of the fastest emerging application in 3G Systems. It also goes on to describe the key specifications of streaming video. The authors describe MPEG-4 as the protocol of choice due to its widespread standardization in the mobile communications industry. The paper gives a detailed introduction of the MPEG-4 Standard. A typical MPEG-4 encoder generates three types of frames. The Intra-frames(I) which contain information from the encoded still image. Prediction frames (P) are directional frames generated from previous P and I frames, and B frames are generated from preceding and following I and P frames. Each video sequence is composed of a repeating sequences of these frames termed as Groups of Pictures (GOP's).

2. THE INTEGRATED TOOL ENVIRONMENT

This section describes the Video Quality Evaluation tool used. The Integrated tool Environment used for the paper in review utilized the Video Acquisition, encoding, designed by Klaue et al. [3]. The tool is a complete end to end frame work to perform Video Quality Assessment over any kind of an IP network. The tool also has a flexibility of choice between a real network environment or a simulator [4]. Whether using a real transport environment or a simulated one, it visualizes the whole framework from the recording/playback at the sender to encoding into MPEG-4, packetization, transmission over the lossy network, jitter reduction by the playout buffer, decoding and finally displaying it at the receiver.

For evaluation of video quality, it is imperative to compare the received video with the transmitted video, but it is impractical to transmit the whole received video back to the sender, which can be really large in size. This problem is solved by using video traces [5]. Video Traces have been declared the most suitable or applicable method to perform a quality estimation between two points over a network [6]. Instead of using real bit streams, which contain all the information carrying bits, traces only give the number of bits used for the encoding of the individual video frames. Let $\mathbf{X_n}$, (where n = 1, . . . , N) denote the size of the frame in bits, then the encoded (compressed) video frame n, whereby N is the total number of frames in the video. A video trace is composed of rows of text, where each row is typically comprised of the frame index (number) n, the frame type (I, P, or B), the frame size in bytes and the time offset of the frame. This time offset of each frame is then used as reference points to calculate the packet loss or delay encountered while traveling through the wireless network. Depending upon the nature of the application, this framework can either use standard video sequences that have predefined trace files associated with them. Online repositories like [7] contain huge sets of standard video sequences with trace files, that can be used for testing and optimization purposes. Fig.2 shows a block diagram of the framework used to assess the video quality at the receiver using trace files from the transmitted video sequence. This data is then used in conjunction with packet dumps from the transmitter and from the receiver over a feedback loop. The utility used to capture packets in this case is tcpdump® [8], although any relevant packet capture tool can be used.

The paper assumes that the packet transmission protocol used is UDP (User Datagram Protocol). UDP is a connectionless protocol that transports packets without retransmissions in a 'fire and forget' manner, making it ideal for real time applications. Streaming media, real-time multiplayer games and voice over IP (VoIP) are typical examples. Lacking reliability, UDP applications must

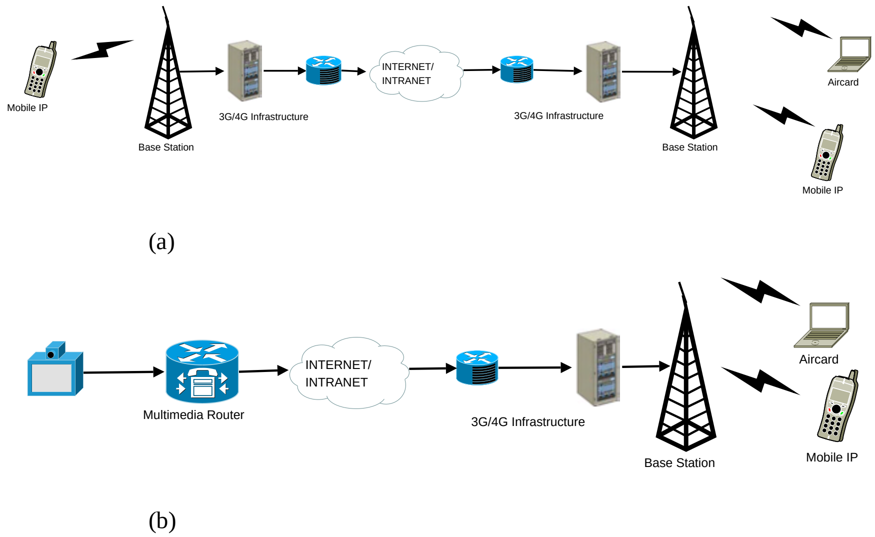

Fig. 1. Typical 3G/4G Wireless Architecture. (a) Mobile to Mobile (b) Land to mobile

generally be willing to accept some loss, errors or duplication. UDP is resides on top of the IP protocol which provides it with unique Packet Identifiers and checksum.

A video frame is relatively bigger than the maximum allowable packet size or Maximum Transfer Unit (MTU) for a UDP packet. 3G wireless networks designers are allowed to choose MTU values ranging from small values (such as 576 bytes) to a large value up to 1500 bytes for IP packets on Ethernet). The paper also points out that it is likely that a Frame size can be much larger than the MTU. This would mean that one frame will be segmented into multiple packets that are transmitted separately. We define the percentage of Packet Loss as:

$$PL_T = (1 - \frac{n_{T_{recv}}}{n_{T_{sent}}}) \times 100 \quad (1)$$

where T: Particular Type of data in packet (header, I, P, B, S). (contrary to the relationship presented in [3], which actually is Packet success)

The paper argues that for frames that are contained within one packet, the packet loss will result in frame loss, but for a larger frame, it is possible that the frame may still be partially recovered by the decoder, hence not resulting in frame loss, but resulting in degradation in quality after passing through the decoder.

The paper also tends to ignore an important fact associated with Decoders designed to deal with the UDP protocol. A major drawback of using the UDP/IP protocol is that single bit errors would result in the whole packet getting discarded due to the checksum failure of the packet. This in turn results in the decoder to drop the whole frame and copy the parameters from the previous frame into the current frame location. This results in peculiar artifacts causing the video quality of that frame to get degraded. This error is subsequently propagated to downstream frames in case the affected frame was an I or P frame. This indicates that the error is not contained within the affected frame but is rather propagated to all the frames in the GOP downstream of the bit error, as shown in 2. This substantial loss of quality emanating from one bit error can be addressed by using a new protocol specifically designed to deal with audio/video transmission error prone networks called UDP-lite [9]. This protocol is designed to avoid checksum calculations on specific portions of the packet payload. This feature will result in far better perceived quality at the receiver for the same bit error rates (BER). So once UDP-lite is utilized as a protocol of choice video transmission protocol, many exciting bit error based Error concealment algorithms can be utilized to greatly improve the overall video quality at the receiver. It is important to note that the preceding UDP and RTP based protocols are so extensively utilized in current network infrastructures that making a protocol change will be an enormous undertaking, both technically and monetarily.

3. VIDEO QUALITY ASSESSMENT

Video Quality assessment of Trace based data can be categorized as a kind of a Full Reference based method. Although there is an assumption that the only parameters that are causing the quality degradation are due to transmission errors over the wireless network. Then the only required features needed at the source side to reconstruct the received video are the information regarding Packet Loss $\mathbf{PL_T}$ and frame jitter $\mathbf{j_F}$. In this section we examine the results from several Objective Image/Video Quality Assessment methodologies. Several standard video sequences [7] were evaluated using this Framework. Most of the papers that have studied Video Quality issues over networks have described PSNR as their standard Objective Video Quality assessment methodology based on its apparent simplicity and well cited findings by the final report from VQEG (in March 2000) on the validation of objective models of video quality assessment [10]. The report declared that, ”No one objective model outperforms the other in all cases”. To validate or disprove these findings from VQEG, various quality assessment methodologies were evaluated on the same sets of data. These methodologies are listed below:

- Mean Square Error (MSE).
- Peak Signal to Noise Ratio (PSNR).

- DCT based Video Quality Estimation [11].
- Mean Spatial Doman Structural Similarity Index (MSSIM) [12].

The MSE and its derivative PSNR are conventional metrics to compare any two images. MSE measures the difference between the original and distorted pixels. PSNR is an logarithmic representation of the inverse of this measure. Compared to other objective measures, PSNR is easy to compute and well understood by most researchers. However the correlation with subjective measure is poor as seen in 2. The subtle differences between degradations of different intensities are not properly reflected using PSNR. Although no thorough Subjective Quality Analysis has been done in this paper to prove this point.

The DCT based Video Quality Estimation also proves to be a good measure for video quality estimation, but our tests reveal that it does not prove to be a good choice when quantifying video sequences that were severely degraded.

The MSSIM proved to be a metric that was closest to a human perception of the received video sequence. This method utilizes structural distortion as an estimate of perceived visual distortion. Where as most other proposed approaches are error sensitivity-based methods [13].

4. CONCLUSION

The first part of this paper reviews a practical evaluation framework that is being proposed for MPEG-4 Video Quality Estimation in a typical cellular wireless network. The framework uses video trace information from the original video sequence to evaluate the video quality degradation at the receiver side. This methodology provides a practical approach to Video Quality Assessment of MPEG-4 Video over 3G broadband wireless networks. The advantages and limitations of this approach were then discussed. The second part of this paper utilizes this framework to study the existing and most recent Objective Image quality assessment algorithms. We found there were some limitations associated with error sensitivity based algorithms like MSE and PSNR, and DCT block error based methods. The structural similarity based approach (MSSIM) proved to be a better metric for video quality over different levels of degradation. A Framework was developed as part of this research that can be used as a test bed to test novel VQA approaches. It can also be used for new Encoding schemes, compression algorithms, Motion Estimation methods and error resilience algorithms etc. This test bed can be used to generate training data for Iterative schemes or ANNs, to be used by the Wireless Vendors to Quantify the Video Quality Degradation over their infrastructure based on key information regarding the Video Transmission. e.g. RF fading, total packet loss, flow control, channel bandwidth used, duration of the transmission, Frame rate, Video resolution (CIF(Common Intermediate Format), QCIF (Quarter Common Intermediate Format) etc.). This methodology can be used by Wireless service providers, to compare real-time traces of their transmitted videos with that of received videos over an error resilient feedback loop. This feedback loop can utilize TCP/IP with retransmissions to ensure error free transmission of its contents. This data will then be used to calculate the perceived video quality at the receiver. This will provide the service providers with real-time video quality analysis over their network, which can help them dramatically improve the quality of their video transmissions and troubleshoot problems more efficiently.

5. REFERENCES

[1] 3GPP2, "cdma2000 High Rate Packet Data Air Interface Specification," *3GPP2 C.S0024-A Veraion 3.0*, Sept. 2006.

[2] G. H. A. Lo and I. Niemegeers, "Evaluation of MPEG-4 Video Streaming over UMTS/WCDMA Dedicated Channels," *Proc. IEEE Int. Conf. Wireless Internet*, vol. 1, pp. 8–10, Jan. 2005.

[3] B. R. J. Klaue and A. Wolisz, "Evalvid - A Framework for Video Transmission and Quality Evaluation," *Proc. 13th Intl Conf on Modeling, Techniques and Tools for Computer Performance Evaluation, Urbana, IL*, 2003.

[4] C.-H. Ke, "An advanced simulation tool-set for video transmission performance evaluation," `http://hpds.ee.ncku.edu.tw/~smallko/ns2/ns2.htm`.

[5] P. Seeling, *Video traces for network performance evaluation : a comprehensive overview and guide on video traces and their utilization in networking research*. Springer, 2007.

[6] M. R. P. Seeling and B. Kulapala, "Network Performance Evaluation using frame size and quality traces of single-layer and two-layer video: A Tutorial," *IEEE Communications Surveys and Tutorials*, vol. 24, no. 4, 2004.

[7] video trace research group at ASU, "YUV Video Sequences," `http://trace.eas.asu.edu/yuv/index.html`.

[8] "TCPDUMP-dump traffic on a network," `http://www.tcpdump.org`.

[9] P. Larzon, Degermark, "UDP Lite for Real-Time Multimedia Applications," *Proceedings of the IEEE International Conference of Communications (ICC)*, 1999.

[10] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000. `http://www.vqeg.org/`.

[11] F. Xiao, "DCT-based Video Quality Evaluation," Final Project for EE392J Stanford Univ. 2000.

[12] Z. Wang, "The SSIM index for image quality assessment," `http://www.cns.nyu.edu/~zwang/files/research/ssim/`.

[13] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment. Synthesis Lectures on Image, Video and Multimedia Processing*. Morgan and Claypool, 2006.

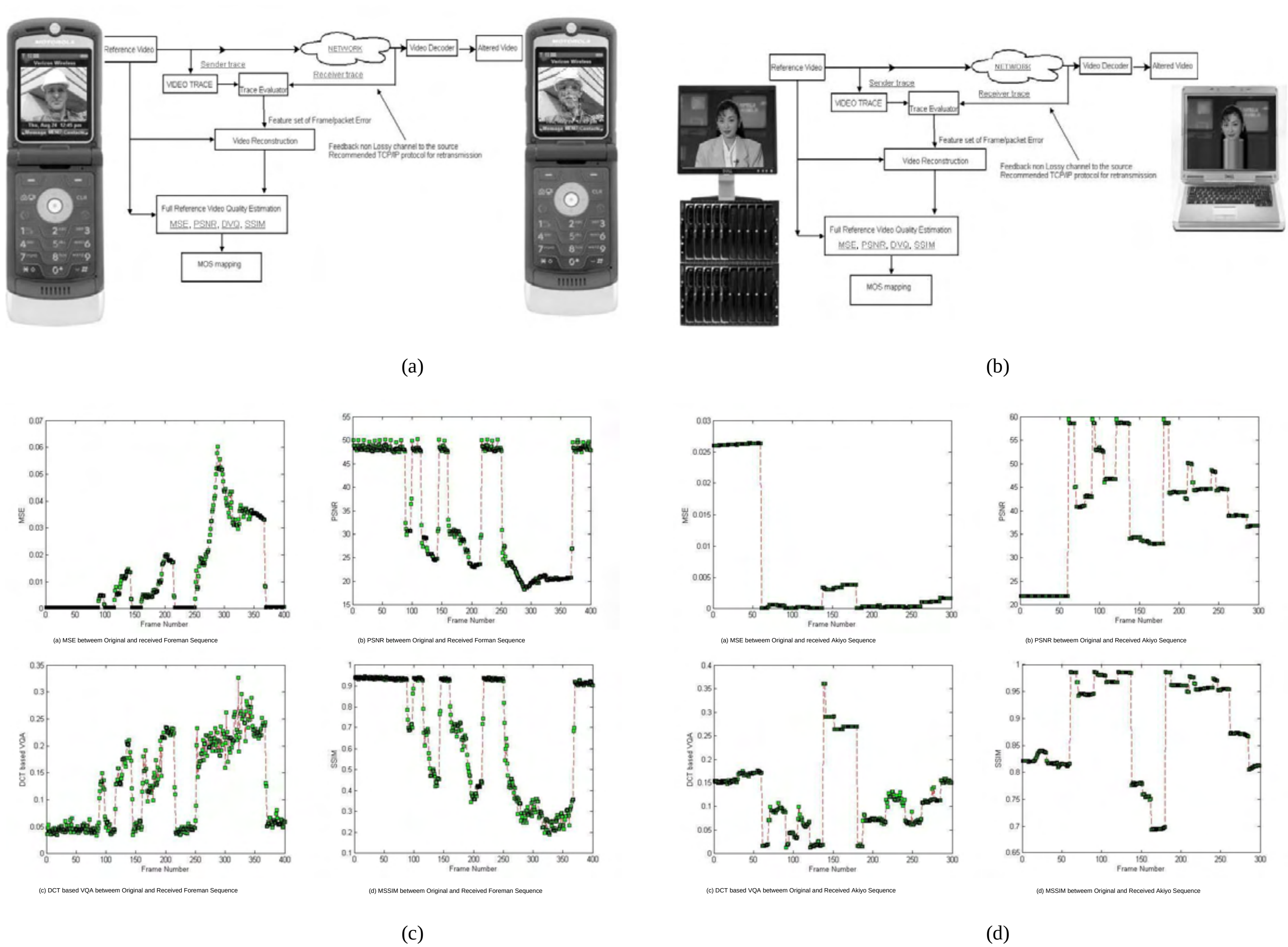

Fig. 2. The Experimental framework for video quality assessment. (a)A Mobile to Mobile example (Foreman, QCIF) (b)A Land to Mobile example (Akiyo, CIF) (c)VQ analysis of the Foreman sequence using MSE, PSNR, DCT based metric and MSSIM (d) VQ analysis of the Akiyo sequence.