

# ERROR CONCEALMENT AND PERFORMANCE EVALUATION OF H.264/AVC VIDEO STREAMS IN A LOSSY WIRELESS ENVIRONMENT

*Muhammad Saleem Koul*

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

## ABSTRACT

The H.264 standard brings the promise of providing real time and streaming multimedia over broadband wireless networks. The current developments in the Fourth Generation Broadband wireless technology, like CDMA-2000 and WiMax [1] provide us with a backbone for such real time multimedia applications. There is however a concern regarding video transmission over RF, i.e. the presence of packet loss and jitter (packet delay-variance). In this paper we analyze the Error Concealment algorithms proposed in the latest H.264 standard. Using the JM13.2, we compare the quality of the decoded video against the original. 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 a wireless environment. We also modify the strategy in the area of Video Quality Assessment by doing a comparative analysis of several major Video Quality assessment methodologies, like PSNR, SSIM and DVQ metric. Comparative conclusions are drawn based on the effect of the amount of packet loss or jitter on a particular Error Concealment technique at the decoder.

## 1. INTRODUCTION

Video transmission over lossy and error prone communication networks is prone to errors introduced during transmission. Transmitting compressed video streams like H.264, makes the problem even worse. There are several stages at which the error introduced during transmission can be corrected or reduced [4]. Video conferencing and live interactive video feeds over wireless networks are more error prone and less flexible towards retransmission based algorithms like automatic repeat request (ARQ) over TCP/IP. A lot of research has gone into devising forward-error correction (FEC) algorithms for recovering lost data segments [5]. FEC algorithms are designed with the requirement that the encoding servers send extra information along with the original video data. With proper amount of redundant data included in the transmitted packets the FEC can mitigate the impact of packet loss in the quality of the video, thus improving the performance of streaming video over error prone networks. These algorithms are not always applicable as they result in increased overhead in terms of bitrate, and usually require a change in the encoding standard.

[6, 7] propose a temporal error concealment algorithms to conceal macroblocks based on the recovered motion vectors using surrounding available macroblocks. However, the algorithms deal with situations where only a portion of the video frame is missing. In a realistic 3G wireless environment, the most common video standard is QCIF, which has a frame size comparable to 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[1]. So it is very likely that a frame will be encapsulated into one transmission packet. This would mean that one packet loss would usually result in one frame loss. Error concealment becomes more difficult in this case, because there are no spatial neighbors available that can help reconstruct the lost frame. In [8–10] several approaches have been proposed for error concealment in frame loss.

## 2. A PACKET LOSS MODEL

Here we discuss a general packet loss model that explains the quality degradation of MPEG-4 due to packet loss [11]. The MPEG-4 compression standard achieves high compression ratios by exploiting spatial and temporal redundancies in consecutive video frames. 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 or I frames, and B frames are generated from preceding and following I or P frames. Each video sequence is composed of a repeating sequences of these frames termed as Groups of Pictures (GOPs). The use of these redundancies helps achieve higher compression ratios in the video sequences, but makes the video sequence susceptible to error propagation due to inter-frame dependencies. A successful decode of a bit-stream with inter-frame dependencies relies on the successful decoding of the reference I-frame and to a lesser degree the P-frames. In this section we will try to analyze the inter-frame dependencies in MPEG-4, focusing on the effect of packet loss in I-frames, and how it affects overall quality of the video stream. The model is based on the assumption that the packet loss will result in the degradation of quality of the video stream at the receiver, and the packet loss will result in the frame not being decodable at the receiver. We can define packet success as a ratio of received vs. transmitted packets. Conversely packet loss can then be defined as the following relationship:

$$p = 1 - \frac{n_{T_{recv}}}{n_{T_{sent}}} \quad (1)$$

where T: Particular Type of data in packet (header, I, P, B).

The current decoder implementations are designed to drop a frame when packet loss occurs in it. Another assumption is that the size of an average I-frame is 5 times or greater than the size of an average P-frame, considering temporal similarities and motion vector based compensation utilized in predictive frames. Experimental results have revealed that the measured results of the

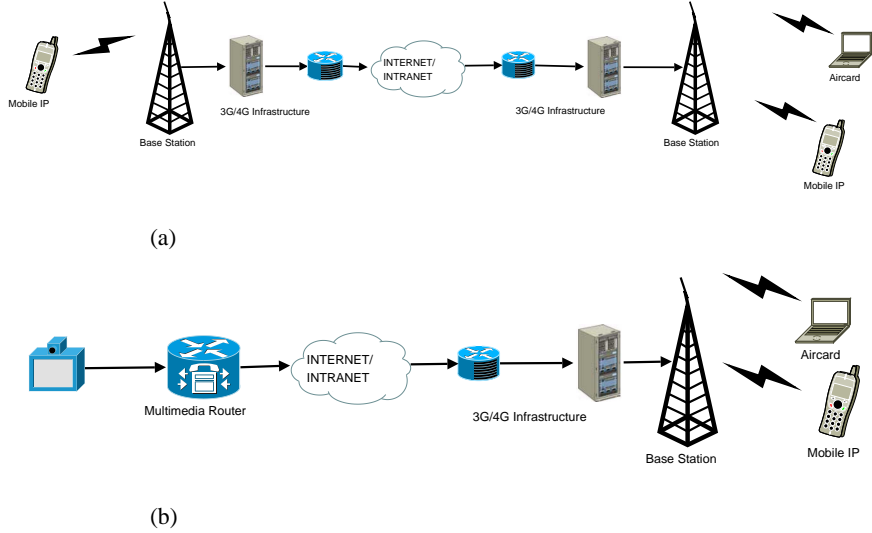


Fig. 1. Typical 3G/4G Wireless Architecture. (a) Mobile to Mobile (b) Land to mobile

resulting frame rates as a function of packet loss rate can be approximated by the equation:

$$f(p) = \alpha(1 - p)^c \quad (2)$$

where  $\alpha$  and  $c$  are constants. The function  $f(p)$  can be considered a Bernoulli random variable, that is directly proportional to the success rate of a video frame. Conversely, we can define a conditional probability  $P$ , for each frame type  $f_i$ , where  $P(\bar{F}|f_i)$  is the probability that a frame of type  $f_i$  was not decoded successfully at the receiver.

$$P(\bar{F}|I) = 1 - (1 - p)^{S_I} \quad (3)$$

where  $S_I$  is the number of packets on average in an I-frame, and  $p$  is packet loss rate. The conditional probabilities for P-frames require the understanding of inter-frame dependencies. The successful decoding of a P-frame depends on all I and P-frames that precede it in the GOP.

$$P(\bar{F}|P) = \frac{1}{N_P} \sum_{k=1}^{N_P} (1 - (1 - p)^{S_I + kS_P}) \quad (4)$$

where  $S_P$  is the number of packets on average in a P-frame, and  $N_P$  is the number of frames in the GOP. We have not considered the B-frames in our current implementation. The Fig.2 shows a plot of Probability of decode failure of I-frames and P-frames plotted against bit error rates ranging from  $2^{-8}$  to  $2^{-3}$ . The plot shows the inter-frame dependencies between the I and P-frames. The probability of unsuccessful decode of P-frames changes with that of I-frames.

### 3. FRAME LOSS ERROR CONCEALMENT ALGORITHMS

In this project we will be more focused on the error concealment of reference frames, particularly Prediction frames (P). In this section we tested the two implemented error concealment algorithms

in the JM13.2 decoder with several other proposed error concealment methods [12]. A novel error concealment algorithm using deformable surfaces based morphing [?] is proposed and applied to the H.264 decoder. There are several applications of this morphing technique, also known as geometric warping. The algorithm adopted in our application was inspired by [?].

- frame copy algorithm: With the algorithm, each pixel value of the concealed frame is copied from the corresponding pixel of the previous decoded reference frame. While concealing a reference frame, the concealed frame is used for display, and is also placed into the reference picture buffer to be used for decoding subsequent pictures. In case of non reference frame concealment, the lost concealed frame is only used for display. An optional de-blocking filtering process can be applied.
- motion vector copy algorithm: In the algorithm, the motion vectors and reference indices of the co-located blocks in the previously decoded reference frame are copied to the lost frame first. The motion vectors are scaled based on the distance of the reference frame from the concealed frame. Then motion compensation is used to reconstruct the lost frame based on the copied motion information. In motion vector copy, the reference frame can be any frame available in the decoder buffer which carries motion information. So even if an IDR frame is lost, as long as it is not the first frame in the bitstream, it can still be concealed with motion vector copy algorithm by specifying the reference frame available in the decoded buffer, possibly from the previous GOP. An optional de-blocking filtering process can be applied.
- "weighted averaging" algorithm: The simplest and often used method is weighted averaging. Each pixel  $p(i, j)$  of a missing macroblock is interpolated as a linear combination of the nearest pixels in the boundaries.
- "decoding without residuals": If the residuals are lost but

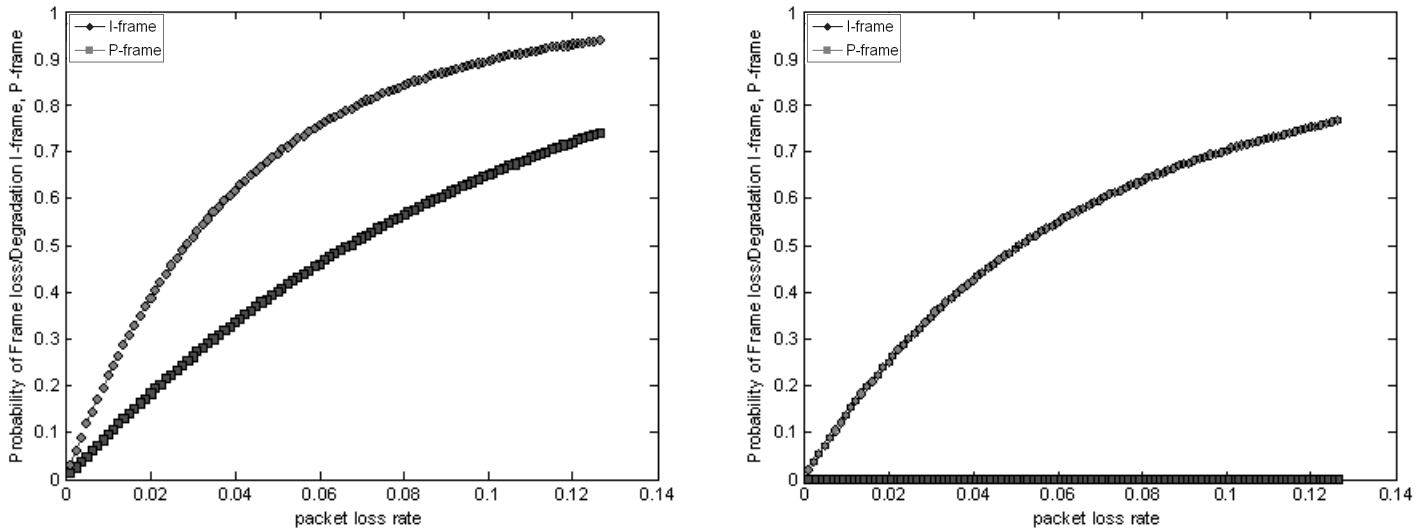


Fig. 2. (a) Probability of Loss of I and P-frames at BER range of  $2^{-8}$  to  $2^{-3}$ . (b) with Probability of loss of I-frames=0

the motion vectors (MV) are correctly received, the simplest is to decode the missing block by setting the missing residuals to zero. This scenario occurs if data partitioning is used for H.264/AVC and motion vectors are better protected than the residuals. Decoding without residuals performs well if the missing residuals were small.

- "geometric morphing based smooting after motion vector copy": We investigate the smoothing properties of image morphing to help conceal errors produced due to loss of multiple blocks. Fig.3 shows that this method presents an improvement in terms of Image quality after applying this algorithm. The Human Visual Sensitivity based method [14] clearly shows the improvement due to this method.

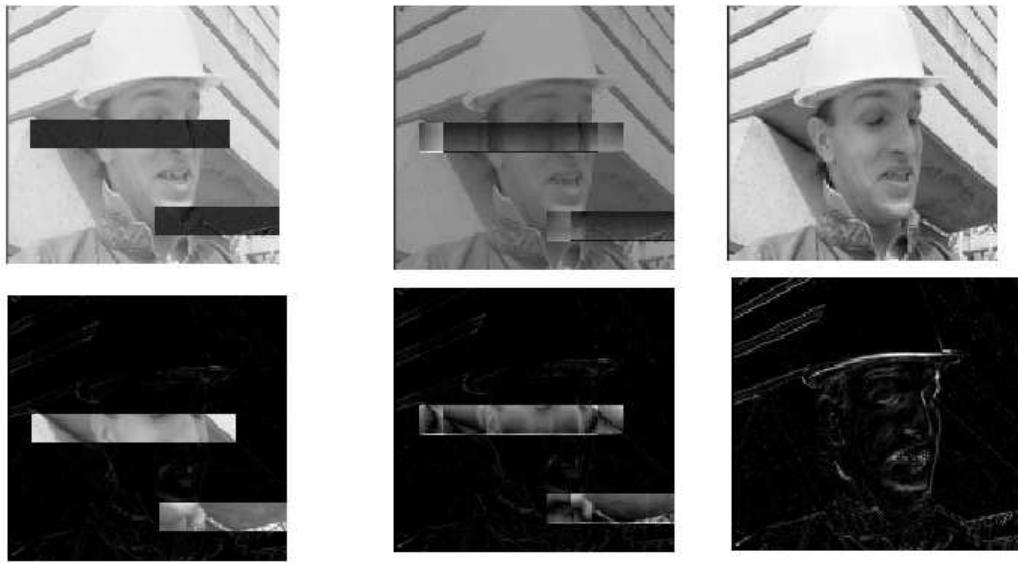
#### 4. THE INTEGRATED TOOL ENVIRONMENT

This section describes the Video Quality Evaluation tool used. The integrated tool environment used for this work utilizes the video acquisition and 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 [15]. 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. Fig.4(a) shows a detailed block diagram of the framework.

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 [16]. Video traces have been declared the most suitable or applicable method to perform a quality estimation between two points over a network [17]. 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  $X_n$ , (where  $n = 1, \dots, N$ ) 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. Web based repositories like [18] contain huge sets of standard video sequences with trace files, that can be used for testing and optimization purposes. Fig.4 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<sup>®</sup> [19], although any relevant packet capture tool can be used.

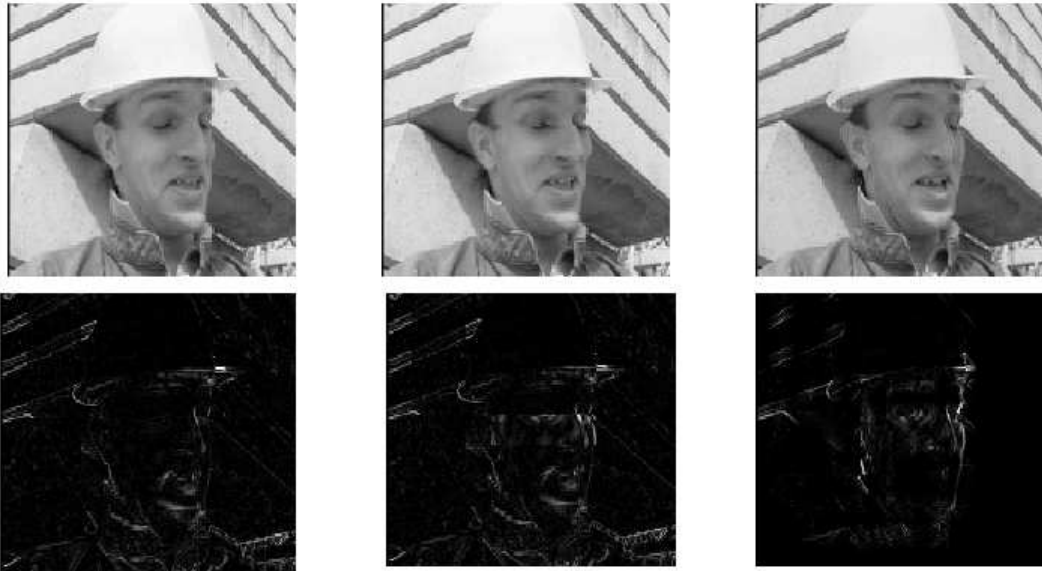
The packet transmission protocol widely used in the 3G wireless networks is UDP (User Datagram Protocol). A major drawback of using the UDP/IP protocol is that single bit errors would result in the whole packet getting discarded due to the checksum failure of the packet or frame. This results in peculiar 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 Fig.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 [20]. 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



(a)

(b)

(c)



(d)

(e)

(f)

**Fig. 3.** (a) (No Error Concealment) MSE=2498, PSNR=14.15 dB, SSIM=0.734 (b) (Weighted averaging) MSE=891.24, PSNR=18.63 dB, SSIM=0.752 (c) (Frame Copy) MSE=123.8, PSNR=27.20 dB, SSIM=0.860 (d) (Decoded without residual) MSE=46.10, PSNR=31.49 dB, SSIM=0.925 (e) (Motion Vector copy without residual) MSE=52.50, PSNR=30.93 dB, SSIM=0.913 (f) (Morph warping from previous MV) MSE=42.51, PSNR=31.85 dB, SSIM=0.928

same bit error rates (BER). 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.

## 5. 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  $PL_T$  and frame jitter  $j_F$ . In this section we examine the results from several Objective Image/Video Quality Assessment methodologies. Several standard video sequences [18] 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 on the validation of objective models of video quality assessment [21]. 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 [22].
- Mean Spatial Doman Structural Similarity Index (MSSIM) [14].

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 Fig.4(b). 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 [23].

## 6. CONCLUSIONS

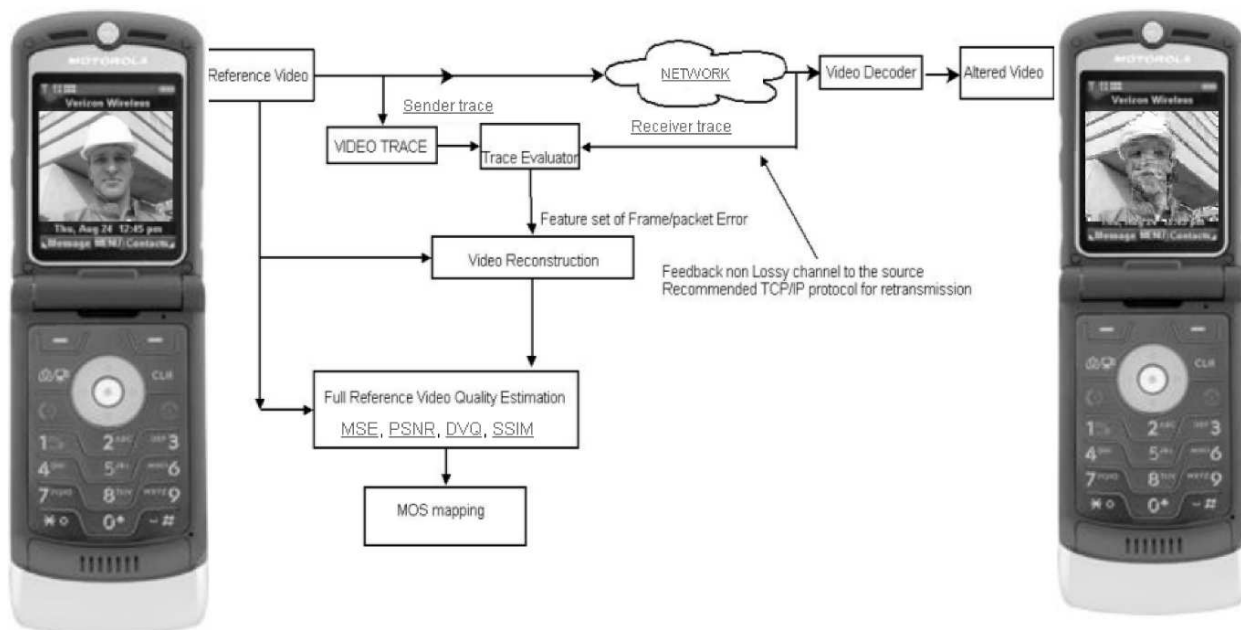
The first part of the proposed research studies the various error concealment techniques applied in the H.264 standard. The intent is to evaluate the performance of these error concealment strategies in an error prone network. Different error concealment algorithms will be implemented using the latest JM13.2. The second part of this paper reviews a practical evaluation framework that is being

proposed for H.264 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 third 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.

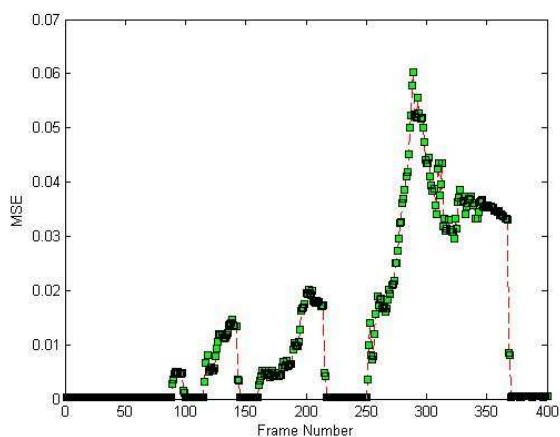
## 7. REFERENCES

- [1] 3GPP2, "cdma2000 High Rate Packet Data Air Interface Specification," *3GPP2 C.S0024-A Veriaion 3.0*, Sept. 2006.
- [2] A. Lo and G. Heijenk 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] J. Klaue and B. Rathke 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] Y. Wang, S. Wenger, J. Wen, and A. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, July 2000.
- [5] S.Kang and D.Loguinov, "Impact of FEC Overhead on Scalable Video Streaming," *NOSSDAV: Network and Operating System Support for Digital Audio and Video.*, June 2005.
- [6] J. Zheng and L. Chau, "A temporal error concealment algorithm for H.264 using lagrange interpolation,," *Proc. of IEEE Int. Symposium on Circuits and Systems*, vol. 2, pp. 133–136, May 2004.
- [7] J. Zheng and L. Chau, "H.264 video communication based refined error concealment schemes,," *IEEE Trans. Consumer Electronics*, vol. 50, pp. 1135–1141, Nov. 2004.
- [8] Y. Chen, J. L. K. Yu, and S. Li, "An error concealment scheme for entire frame loss in video transmission,," *IEEE Trans. Consumer Electronics*, vol. 50, pp. 1135–1141, Nov. 2004.
- [9] P. Baccichet, D. Bagni, A. Chimienti, and L. Pezzoni, "Frame concealment for H.264/AVC decoders,," *IEEE Trans. Consumer Electronics*, vol. 51, pp. 227–233, Feb. 2005.
- [10] Z. Wu and J. M. Boyce, "An error concealment scheme for entire frame losses based on h.264/avc,," *ISCAS*, pp. 4–8, Apr. 2006.
- [11] N. Feamster and H. Balakrishnan, "Packet Loss Recovery for Streaming Video," *12th International Packet Video Workshop*, Apr. 2002.
- [12] I. C. Todoli, *Performance of Error Concealment Methods for Wireless Video*. PhD thesis, Vienna University of Technology, 2007.

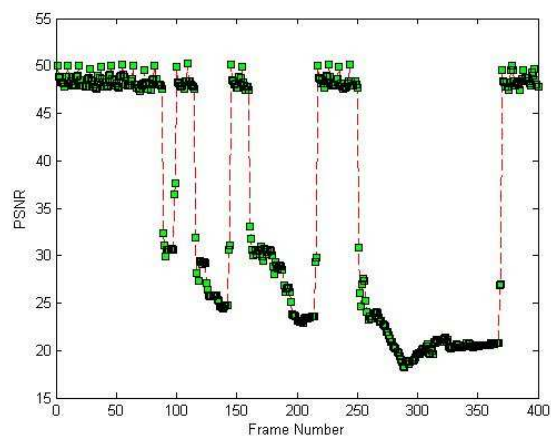
- [13] Y. Wang, Y. Chen, and A. Amini, "Fast LV motion estimation using subspace approximation techniques,," *J. of Medical Imaging*, vol. 20, pp. 499–513, June 2001.
- [14] Z. Wang, "The SSIM index for image quality assessment," <http://www.cns.nyu.edu/~zwang/files/research/ssim/>.
- [15] C.-H. Ke, "An advanced simulation tool-set for video transmission performance evaluation," <http://hpds.ee.ncku.edu.tw/~smallko/ns2/ns2.htm>.
- [16] P. Seeling, *Video traces for network performance evaluation : a comprehensive overview and guide on video traces and their utilization in networking research*. Springer, 2007.
- [17] P. Seeling, M. Reisslein, 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.
- [18] video trace research group at ASU, "YUV Video Sequences," <http://trace.eas.asu.edu/yuv/index.html>.
- [19] "TCPDUMP-dump traffic on a network," <http://www.tcpdump.org>.
- [20] Larzon, Degermark, and Pink, "UDP Lite for Real-Time Multimedia Applications," *Proceedings of the IEEE International Conference of Communications (ICC)*, 1999.
- [21] 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/>.
- [22] F. Xiao, "DCT-based Video Quality Evaluation," Final Project for EE392J Stanford Univ. 2000.
- [23] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment. Synthesis Lectures on Image, Video and Multimedia Processing*. Morgan and Claypool, 2006.



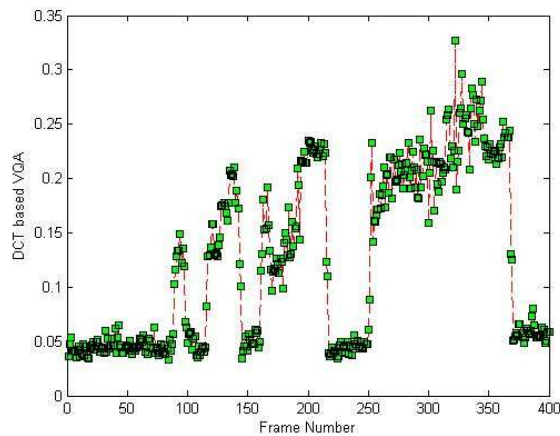
(a)



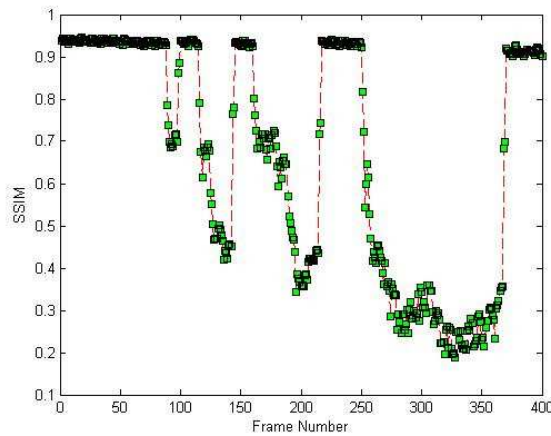
(a) MSE between Original and received Foreman Sequence



(b) PSNR between Original and Received Foreman Sequence



(c) DCT based VQA between Original and Received Foreman Sequence



(d) MSSIM between Original and Received Foreman Sequence

(b)

**Fig. 4.** The Experimental framework for video quality assessment. (a) A Mobile to Mobile example (Foreman, QCIF) (b) VQ analysis of the Foreman sequence using MSE, PSNR, DCT based metric and MSSIM