Advances in Engineering Science
Sect. C (3), July-September 2008, PP 9-16

ERROR CONCEALMENT TECHNIQUES IN H.264/AVC FOR WIRELESS VIDEO TRANSMISSION IN MOBILE NETWORKS

Vineeth Shetty Kolkeri*, M. S. Koul, J. H. Lee and K. R. Rao

IEEE Fellow, Department of Electrical Engineering, University of Texas at Arlington

ABSTRACT

Nowadays audio-visual and multimedia services are seen as important sources of the traffic within mobile networks. An important characteristic of such networks is that they cannot provide a guaranteed quality of service (QoS) due to issues like interfering traffic, signal to noise ratio fluctuations etc. One of the limitations within the mobile networks is the low transmission bit rate which demands the reduction of the video resolution and an efficient video compression technique such as H.264/AVC [8] standard, which provides a high compression gain. 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 H.264 standard. Then this strategy is modified to suit our application in a wireless environment. Subjective quality of the error concealed images is analyzed based on PSNR, MSE and SSIM. 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.

Index Terms : *Error Concealment, H.264/AVC, video coding standards, Wireless video transmission.*

I. INTRODUCTION

Video transmission over lossy Wireless 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 errors introduced during transmission can be corrected or reduced [1]. 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 [12]. A lot of research has gone into devising forward-error correction (FEC) algorithms for recovering lost data segments [2]. 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 [11]. 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.

The H.264/AVC [8] standard consists of two layers, the video coding layer (VCL) and the network abstraction layer (NAL). The VCL specifies an efficient representation for the decoded video data. It is designed to be as network independent as possible. The coded video data is organized into NAL units, each of which is a packet that contains an integer number of bytes. The first byte of each NAL unit is a header byte that contains an indication of the type of data in the NAL unit, and the remaining bytes contain payload data of the type indicated by the header [2]. In the video transmission, the order in which the NAL units have to be sent is constant. The first NAL unit to be sent is the sequence parameter set (SPS) followed by the picture parameter set (PPS). Both SPS and PPS include some parameters which have been set in the encoder configuration for all pictures in the video sequence, for example:

---

*vineeth.kolkeri@mavs.uta.edu

entropy coding mode flag, number of reference index, weighted prediction flag, picture width in MB, picture height in MB and number of reference frames.

The next NAL unit is the Instantaneous Decoder Refresh (IDR). After receiving a NAL unit of this type all the buffers have to be deleted. An IDR frame may only contain I slice without data partitioning. IDR frames are usually sent at the start of the video sequence. For the streaming video services over the mobile technologies, the IP packet switched communications is of major interest, which uses real time transport protocol (RTP). Each NAL unit regardless of its type is encapsulated in the RTP/UDP/IP [12] packet by adding header information of each protocol to the NAL unit. IP header is 20 or 40 bytes long, depending on the protocol version and contains the information about the source and destination IP address. UDP header is 8 bytes long and contains the CRC and length of the encapsulated packet. RTP header is 12 bytes long and contains sequence number and time stamps.

II. PACKET LOSS MODEL

Here we discuss a general packet loss model that explains the quality degradation of MPEG-4 due to packet loss [3]. 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. frames P are predicted frames generated from previous P or I frames, and B frames are generated from proceeding and following I or P frames. Each video sequence is composed of 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 packets nT_{recv} vs. transmitted packets nT_{sent}. Conversely packet loss p can then be defined as the following relationship:

$$P = 1 - \frac{nT_{recv}}{nT_{sent}} \qquad (1)$$

where T: Particular type of data in a 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 number of bits needed for coding an I frame is about 5 times the number of bits needed for coding a P frame, considering temporal similarities and motion vector based compensation utilized in predictive frames. Experimental results have revealed that the measured results of the resulting frame rates as a function of packet loss rate can be approximated by the equation:

$$f(\mathrm{p}) = \alpha\,(1-\mathrm{p})^{\mathrm{c}} \qquad (2)$$

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 I (Intra-Frame), where P(F|I) is the probability that a frame of type I (Intra-Frame) was not decoded successfully at the receiver.

$$\mathrm{P(F|I)} = 1 - (1-\mathrm{p})^{S_{\mathrm{I}}} \qquad (3)$$

where *S*I is the number of packets on the 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.

$$P(F|I) = \frac{1}{N_P}\sum_{k=1}^{N_P}\left(1-(1-p)^{S_I+ksp}\right) \qquad (4)$$

where *S*P is the number of packets on the 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.

III. ERROR CONCEALMENT ALGORITHMS

The loss of transmitted data packets influences the quality of the received video. This problem is caused by the band limited channel used by the mobile communication networks. Since the real time transmission of video stream limits the channel delay, it is not possible to retransmit all erroneous or lost packets. Therefore there is a need for post processing methods, which try to reduce the visual artifacts caused by bit stream error after locating the missing or defect parts of video data. Error concealment methods which shall be implemented on the receiver side restore the missing and corrupt video content by using the previously decoded video data. The error concealment benefits from the spatial and temporal correlations between the video blocks within a frame or more than one frame within the video sequence. Therefore the error concealment methods are implemented in the spatial domain and in the time domain. The spatial domain based error concealment uses the video information from the neighboring blocks to restore the missing pixels within a specified area. The time domain based error concealment uses the video information from the blocks lying in the previous and following frames to restore the missing pixels within a specified area.

In this paper we are more focused on the error concealment of reference frames, particularly prediction frames (P). We tested the two implemented error concealment algorithms in the JM13.2 [13] decoder with several other proposed error concealment methods. A novel error concealment algorithm using deformable surfaces based on morphing is proposed and applied to the H.264 decoder. There are several applications of this morphing technique, also known as geometric warping.

- Frame copy algorithm: In this 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 with motion estimation in combination with weighted averaging algorithm (proposed): In this 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 of a missing macroblock is interpolated as a linear combination of the nearest pixels in the boundaries.

IV. 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. The only required features needed at the source side to reconstruct the received video are the information regarding packet loss and frame jitter. In this section we examine the results from several Objective Image/Video Quality Assessment methodologies. Several standard video sequences [4] were evaluated using this framework [9]. 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 [5]. 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).
- Structural Similarity Index (SSIM) [6].

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 a 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 both MSE and PSNR do not correlate well with the subjective quality of the reconstructed images. The subtle differences between degradations of different intensities are not properly reflected using PSNR. The SSIM 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, whereas most other proposed approaches are error sensitivity based methods [7].

V. EXPERIMENTAL RESULTS

Table 1 : Results for the frame copy algorithm

FOREMAN SEQUENCE (QCIF) FRAME	MSE	PSNR in dB	SSIM
I	12.08661	37.30776	0.96522
P	13.95277	36.6842	0.9626
P	48.85941	31.24132	0.91529
P	48.48063	31.27512	0.92303
P	46.28512	31.47639	0.93049
P	45.00639	31.59806	0.93579
P	44.962	31.60235	0.93632
P	38.08712	32.32302	0.94188
P	34.66	32.73252	0.94407
P	32.84486	32.96613	0.94518
P	32.6877	32.98696	0.94467
P	31.70009	33.1202	0.94771
P	31.16525	33.1941	0.94674
P	30.904	33.23066	0.94634
P	30.99349	33.2181	0.94605
I	12.83641	37.04637	0.96451
P	13.05753	36.97219	0.96349
P	12.97929	36.99829	0.96342
P	12.80007	37.05868	0.96295
P	12.70391	37.09143	0.96313
AVERAGE	29.35263	35.45433	0.94744

Table 2 : Results for the proposed algorithm

FOREMAN SEQUENCE (QCIF) FRAME	MSE	PSNR in dB	SSIM
I	12.08661	37.30776	0.96522
P	13.95277	36.6842	0.9626
P	13.89938	36.70085	0.96128
P	13.88774	36.70449	0.96223

P	53.4302	30.85294	0.9245
P	51.85097	30.98323	0.92781
P	50.39078	31.10729	0.92989
P	51.7794	30.98923	0.9289
P	51.92803	30.97679	0.92784
P	51.55366	31.00821	0.9269
P	52.5406	30.92585	0.92732
P	51.68616	30.99706	0.93093
P	49.52837	31.18226	0.93239
P	49.69014	31.1681	0.93217
P	47.64197	31.35091	0.93337
I	12.83641	37.04637	0.96451
P	13.05753	36.97219	0.96349
P	12.97929	36.99829	0.96342
P	12.80007	37.05868	0.96295
P	12.70391	37.09143	0.96313
AVERAGE	34.01120	32.81458	0.95444

GoP Level: Due to the temporal and spatial predictions of the images, the image distortion caused by an erroneous MB is not restricted to that MB. Since MBs are spatially and/or temporally dependent of neighboring MBs, the errors can also propagate in time (in following frames) and in space (the same frame). Error propagation represents a problem for error concealment because if the error concealed picture differs from the original picture, the error will propagate until the next I frame occurs i.e., until the beginning of the next GoP. If we use more frames per GoP, we can compress better, but the error can propagate over more frames.

In our algorithm the GoP level is taken to be 15 in which it has a sequence of first I frame followed by 14 P frames. The error occurred in the second frame and it propagates to the start of next GoP. Figure 2 show the format of frame which is divided into different macroblock sizes depending on the information present in the frame. If in a macroblock it contains more information then it is divided into smaller sizes. Figure 3 show numbers of bits present in I and P frames in a video sequence. In our simulation we have encoded 20 frames of Foreman sequence (QCIF – Quarter Common Intermediate Format).

Table 3 : Average values of the video quality metric for different video sequences

SEQUENCE	Average MSE	Average PSNR in dB	Average SSIM
Foreman (Copy)	29.35263	35.45433	0.94744
Foreman (Proposed)	34.01120	32.81458	0.95444
Akiyo (Copy)	8.68558	38.74282	0.97244
Akiyo (Proposed)	8.6870	38.750	0.97320
City (Copy)	34.45604	38.74282	0.93090
City (Proposed)	34.475	32.75815	0.93823

(a) Original Frame (b) Damaged Frame

(c) Copy Algorithm (d) Proposed

Figure 1 : Comparison of foreman sequence with different algorithms for isolated block error.

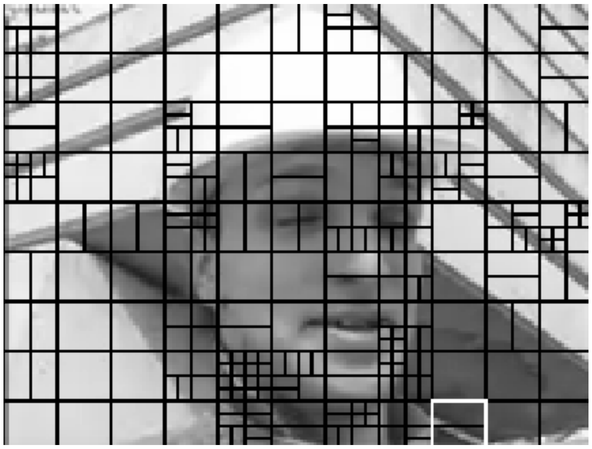
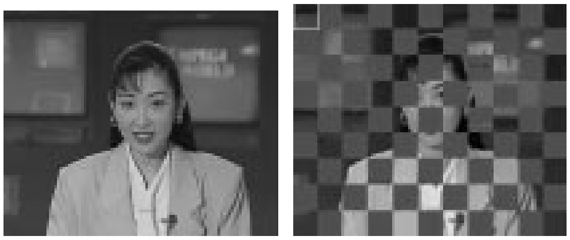


Figure 2 : Foreman frame divided into multiple block sizes of 16 x 16, 16 x 8, 8 x 16, 8 x 8, 4 x 8, 8 x 4, and 4 x 4 to represent coding profiles

Figure 3 : No. of Bits vs. No. of Frames Encoded (Graph shows the size of the different I and P frames obtained after encoding 20 frames of the foreman QCIF video sequence. Green line shows the average values of the bit lost when it is passed through the lossy algorithm after encoding in a video sequence. The results show that frame copy algorithm works better in a video sequence which has very less motion between the frames for concealing the error. Motion vector averaging algorithm works better in cases where there is a high motion between the successive video frames).



(a) Original Frame (b) Damaged Frame

(c) Proposed

Figure 4 : Comparison of Akiyo sequence with different algorithms for isolated block error.

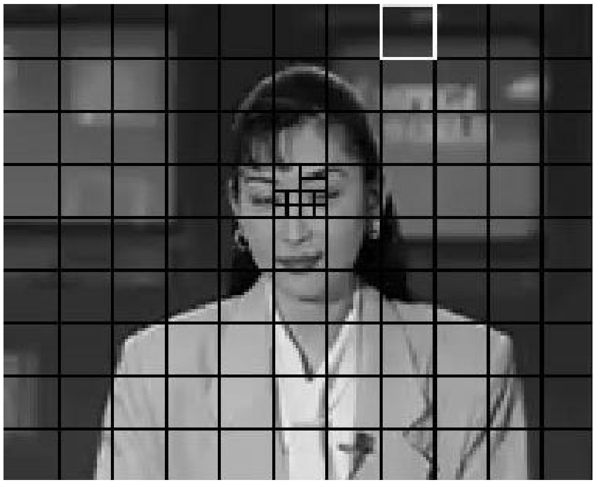
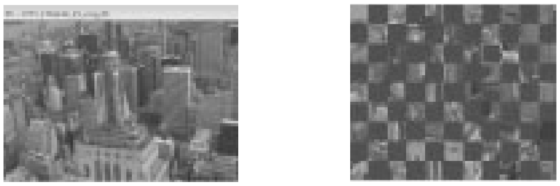


Figure 5 : Akiyo frame divided into multiple block sizes of 16 x 16, 16 x 8, 8 x 16, 8 x 8, 4 x 8, 8 x 4, and 4 x 4 to represent coding profiles



(a) Original Frame (b) Damaged Frame

(c) Proposed

Figure 6 : Comparison of city sequence with different algorithms for isolated block error.

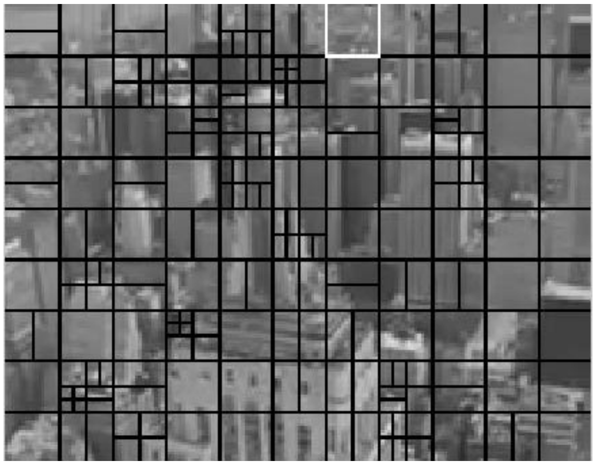


Figure 7 : City frame divided into multiple block sizes of 16 x 16, 16 x 8, 8 x 16, 8 x 8, 4 x 8, 8 x 4, and 4 x 4 to represent coding profiles.

VI. 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 are implemented using the latest JM13.2 [13]. 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 are 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 are some limitations associated with error sensitivity based algorithms like MSE and PSNR, and DCT block error based methods. The structural similarity based approach (SSIM) proved to be a better metric for video quality over different levels of degradation.

REFERENCES

1. Y. Wang, et al, “Error resilient video coding techniques,” IEEE Signal Processing Magazine, vol. 17, July 2000, pp. 61-82.
2. S.Kang and D.Loguinov, “Impact of FEC Overhead on Scalable Video Streaming,” NOSSDAV: Network and Operating.
3. N. Feamster and H. Balakrishnan, “Packet Loss Recovery for Streaming Video,” 12th International Packet Video Workshop, Apr. 2002.
4. Video trace research group at ASU, “YUV Video Sequences,” http://trace.eas.asu.edu/yuv/index.html.

5. 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/.
6. Z. Wang, et al, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. Image Processing, vol. 13, April 2004, pp.600-612.
7. Z. Wang and A. C. Bovik, Modern Image Quality Assessment. Synthesis Lectures on Image, Video and Multimedia Processing. Morgan and Claypool, 2006.
8. Soon-kak Kwon, A. Tamhankar and K.R. Rao, Overview of H.264 / MPEG-4 Part 10", J. Visual Communication and Image Representation, vol. 17, April 2006, pp.186-216.
9. P.Salama, N. Shroff and E. J. Delp, "Error Concealment in Encoded Video Streams", Proc. IEEE ICIP, vol. 1, 1995, pp. 9-12.
10. M. Ghanbari and V. Seferidis, "Cell-Loss Concealment in ATM Video Codecs", IEEE Trans. Circuits and Systems for Video Technology, vol. 3, June 1993, pp. 238-247.
11. T. Stockhammer, M. M. Hannuksela and T. Wiegand, "H.264/AVC in Wireless Environments", IEEE Trans. Circuits and Systems for Video Technology, vol. 13, July 2003, pp. 657- 673.
12. S. Wenger, "H.264/AVC over IP" IEEE Trans. Circuits and Systems for Video Technology, vol. 13, July 2003, pp. 645-656.
13. JM 13.2 software http://iphome.hhi.de/suehring/tml/

■