

An N+1 redundant GOP based FEC algorithm for improving Streaming Video Quality due to Packet Loss and Delay Jitter

M.S. Koul and K.R. Rao
Dept. of Electrical Engineering,
The University of Texas at Arlington, Arlington, Texas 76019
Email: mskoul@uta.edu, rao@uta.edu

Abstract—One of the greatest challenges in deploying streaming video content over wireless is the existence of random packet loss and delay variance (jitter). Due to the nature of real-time transmissions, packet retransmission is not always the preferred viable solution to this problem. Retransmission based error recovery schemes add overhead in terms of time taken for packets retransmissions, which can lead to undesirable packet delays, resulting in video quality degradation. In this paper we propose an error resilient coding scheme using Forward Error Correction (FEC), by encoding every next I-frame as an extra Predictive P-frame as a part of the preceding Group of Pictures (GOP), thus virtually making the GOP size N+1, but adding selective redundancy to the video stream. This scheme for adding redundancy to selected frames in the video transmission makes them less susceptible to errors and thus makes the video stream more robust towards error propagation.

I. INTRODUCTION

Current Multimedia video applications are moving rapidly towards real-time communications and instantaneous transmission of visual information over wireless standards such as CDMA 2000 and EvDO [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. All data communication, including video and audio, is transmitted using the IP protocol (commonly known as the Internet). The IP protocol requires the packets to be of a certain size i.e. Maximum Transfer Unit (MTU), the protocol thus breaks down the frame into multiple constituent packets. Each packet consists of a header along with the payload. UDP (User Datagram Protocol) and TCP/IP require the payload parts of the packet to be checksum protected. So if a single bit flip occurs in the payload part, the whole packet is rendered useless. The current commercial MPEG/H.264 standard decoders are designed to drop the frame whenever it discovers a dropped packet within the frame. The decoder then copies the last frame into the current frame location, on the assumption that temporal information is preserved within consecutive frames.

II. ERROR RESILIENCE USING REDUNDANCY

The choice between error resilience and compression schemes generally requires a trade off, as it is difficult to achieve both strong error resilience and good compression simultaneously. For example, the compression algorithms try

to remove (spatial and temporal) redundancy as much as possible, on the other hand make it more vulnerable to errors. 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.

Many studies investigate the performance of FEC, however, the results on the effectiveness of FEC generally vary and often depend upon several distinct factors. [3] proposes a solution to video data loss by piggy-backing redundant video frames within the transmitted video stream in order to repair lost frames. In their research, they apply an existing IETF (Internet Engineering Task Force) RFC (Request for comments) standard [4] defined for streaming audio. A repetition technique is used to compensate for the loss by repeating the frame that is received immediately before the lost one. If the lost frame is an important frame (I-frame or P-frame, the subsequent frames will be lost or damaged as well, since they are dependent upon the lost frame. By playing the previous frame again, the perceptual quality of video may decrease. The received sequence would show a noticeable jerk or frame freeze followed by a sudden jump from one scene to a totally different one. The redundant frame is kept at a very low quality to avoid overhead. Their tests indicated that by choosing a quality of 25 (on scale of 0-31), significant decrease in frame size was achieved. The ratio of a primary I-frame to the redundant I-frame was kept around 29k to 4.5k respectively.

III. A PACKET LOSS MODEL

Here we discuss a general packet loss model that explains the quality degradation of MPEG-4 in due to packet loss [5]. 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 and I frames, and B frames are generated from preceding and

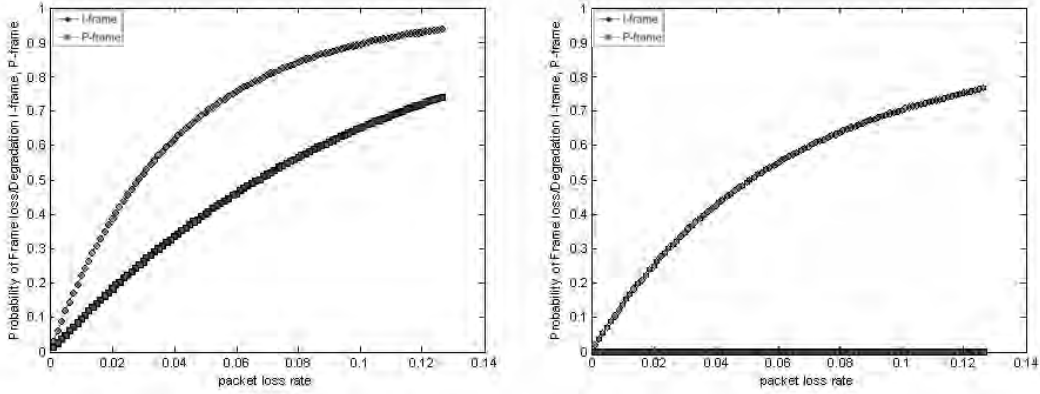


Fig. 1. (a) Probability of Loss of I and P-frames at BER range of 2^{-8} to 2^{-3} . (b) with Probability of loss of I-frames=0

following I and P frames. Each video sequence is composed of a repeating sequences of these frames termed as Groups of Pictures (GOP's). The use of these redundancies helps achieve higher compression rates 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 to become not 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, S).

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 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.

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

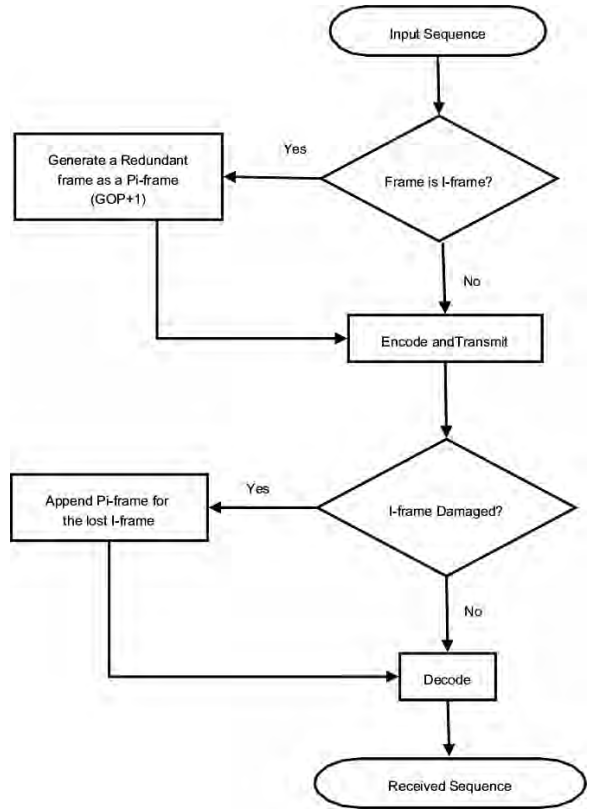


Fig. 2. Proposed flow diagram of the algorithm

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 requires 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.

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

We have not considered the B-frames in our current implementation, so they are not considered. The Fig.1 shows a

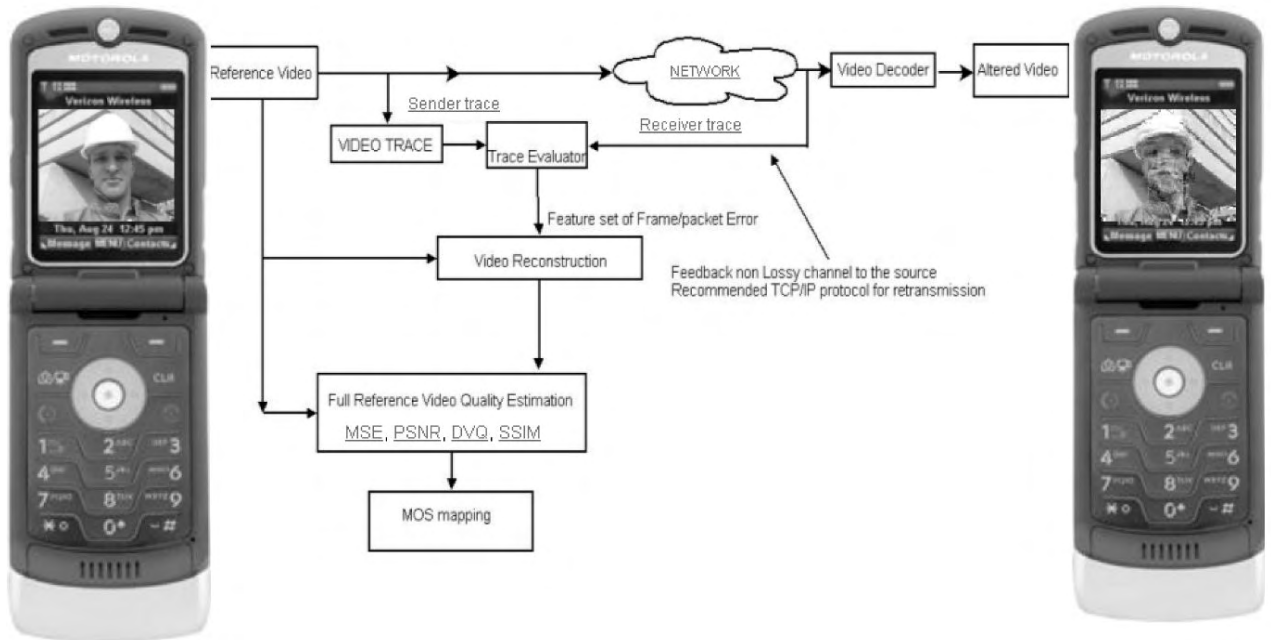


Fig. 3. Framework used for Video Quality Assessment over a Lossy Wireless Network.

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. Thus if we come with an FEC algorithm that would decrease or virtually eliminate the Probability of unsuccessful decode of I-frames. The Probability of unsuccessful decode of P-frames drops significantly. Using this model as an argument, we design an FEC algorithm that specifically increases the resilience of I-frames, which should in turn increase video quality of the whole video stream.

IV. N+1 GOP ALGORITHM

The proposed algorithm shown in Fig.2 is simplistically defined below.

- The Encoder identifies the frame as an I-frame.
- The frame is encoded twice, once as an I-frame and once as a P-frame as part of the previous GOP. We term this frame as frame P_i .
- The frame is transmitted along with the rest of the frames through the bitstream.
- At the receiver, if the receiver detects a damage to the original I-frame, it drops the frame, and instead uses the redundant P_i frame in its place and proceeds to decode the rest of the GOP using that frame as the reference I-frame.
- The ratio of the sizes of I-frame and P-frame is kept around 5:1. This is comparable to the ratio of the original vs. redundant frames proposed by [3].

V. EXPERIMENTAL IMPLEMENTATION AND OBJECTIVE VIDEO QUALITY ANALYSIS

The algorithm was implemented in Matlab[®] and tested in a framework designed to inject and evaluate packet loss and jitter in video bitstreams [6], [7]. Fig.3 shows the detailed implementation of the framework. Some assumptions were applied to the simulations to reduce the computational complexity and simplify the implementation. The following factors were kept constant across all the implementations. The GOP size in this case is limited to 5 frames per GOP with I and P frames only. The GOP pattern is I P P P P. The Macro Block (MB) is set to 16x16 pixel size. The exhaustive search Motion Estimation algorithm was used [8]. Only the Luminance (Gray scale) frames are used. The bandwidth is set to 2 Mbps, and the frame rate is 30 fps. The experimental implementation was done by encoding and then transmitting a bitstream of 10 frames. The same YUV sequence of 10 frames was then encoded using the commercial MPEG-4 decoding algorithm, the one proposed by [3], and our proposed method. All standard encoder parameters like quantization levels, compression, bandwidth were kept constant across all implementations. In each encoded bitstream, the 6th frame, (i.e. the I-frame) was deliberately corrupted, by changing a bit in the payload portion of one of the packets that contained the data for the I-frame. This resulted in a complete UDP packet loss at the receiving end. Fig.4 shows the 10 frames for the 'Mobile' QCIF sequence, followed by the same 10 frames using the commercial MPEG-4 coder, and then our proposed solution. As can be seen the standard decoder drops the frame and copies the previous frame into this location. Although this implementation has the lowest computational complexity,

TABLE I
VIDEO QUALITY ASSESSMENT OF THE FOOTBALL CIF SEQUENCE WITH I-FRAME CORRUPTION AT FRAME 6

| VQ Scheme | Coder | Frame1 | Frame2 | Frame3 | Frame4 | Frame5 | Frame6 | Frame7 | Frame8 | Frame9 | Frame10 |
|-----------|--------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| PSNR | Standard MPEG4 | 46.6837 | 46.6895 | 46.7073 | 46.7135 | 46.7323 | 13.1627 | 13.1643 | 13.3641 | 13.4493 | 13.8195 |
| | Liu et al Solution | 46.6837 | 46.6895 | 46.7073 | 46.7135 | 46.7323 | 17.5249 | 17.3902 | 17.4527 | 17.5564 | 17.686 |
| | Proposed Solution | 46.6837 | 46.6895 | 46.7073 | 46.7135 | 46.7323 | 46.7592 | 46.7675 | 46.7693 | 46.7791 | 46.6855 |
| MSSIM | Standard MPEG4 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.377 | 0.4734 | 0.4824 | 0.4753 | 0.5125 |
| | Liu et al Solution | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.7623 | 0.7793 | 0.7826 | 0.7657 | 0.7817 |
| | Proposed Solution | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

and introduces no overhead, the output sequence downstream of frame 6 starts to degrade rapidly and one can see residual artifacts in all the frames in the GOP. In order to get numerical measures of the effects on video quality in each of the cases, several prevalent approaches such as MSE, PSNR, DCT based Video Quality metric [9] and MSSIM [10] were applied to the original and decoded video sequences. Fig.5 shows the Video Quality assessments by the afore mentioned methods on these sequences. It can be seen that all Video Quality assessment algorithms identify that the standard decoders show a significant decline in video quality at frame 6 (although MSE needs to be scaled to the order of 10^3 to show any apparent change), and the quality stays degraded throughout the remainder of the GOP. The same packet loss was then applied to the proposed encoded sequence and the decoder upon identifying the lost I-frame, used the redundant P_i frame instead for the I-frame. The quality of the received video frame is dramatically improved, and there is no significant video quality degradation reported by our method. Next we encode 10 frames of the 'Football' CIF sequence. This time the sequence was encoded and decoded using the commercial algorithm, our proposed implementation and the solution proposed by [3]. The same 6th frame, i.e. the I-frame was corrupted and hence dropped at the decoders. The original sequence and the decoded sequences from the three decoders are shown in Fig.6. The results show the highest amount of degradation using the current implementation of the MPEG-4 decoders. The results show that [3]'s implementation introduces artifacts due to information lost due to compression. These artifacts then result in further degradation of the predicted frames downstream of this frame. Fig.7 shows the Video Quality assessment of the same 'Football' CIF sequence using PSNR and MSSIM measures. The results from PSNR measures indicate that both the commercial decoder and the previous redundancy solution proposed by [3] results in quality degradation at the lost frame and the error is propagated to the remaining frames in the GOP. Table 1 shows the relative quality metrics of the three type of coding algorithms using the Football video sequence. The results show that our proposed method provided far better error resilience than both the previous redundancy approach and the current decoder implementations.

VI. CONCLUSION

In this paper we try to analyze a Packet Loss Model using Bernoulli random variables. Thus we come up with a model that quantizes dependencies exhibited by P-frames on I-frames

and how decreasing the failure rate of the I-frame can result in decreased failure rate of the P-frames in the GOP. Using this model as a motivating factor for increasing the error resilience of I-frames, we propose a novel but simple method that uses a widely tested area for error resilience and error correction algorithms called FEC (Forward Error Correction). Redundancy is one of the simplest methods of increasing error resilience in a video bitstream. Using selective redundancy, we propose introducing redundancy only to I-frames, by encoding the previous GOP as an N+1 GOP, where N is the number of frames in the GOP. The extra P frame at the end of the previous frame is actually the next I-frame encoded as a P-frame. This effectively decreases the size of the redundant frame but keeps the video quality of the frame virtually unscathed. This algorithm is then tested and compared against commercial and previously proposed implementation of I-frame redundancy. Results are then scrutinized by several Video Quality Metrics. The results show that our implementation shows better error resilience than both the commercial and previously proposed solutions. The results also show that the MSE and PSNR do not give a good objective quality measure between different decoding schemes when the differences are quite subtle.

REFERENCES

- [1] 3GPP2, "cdma2000 High Rate Packet Data Air Interface Specification," 3GPP2 C.S0024-A Verion 3.0, Sept. 2006.
- [2] 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.
- [3] Y. Liu and M. Claypool, "Using Redundancy to Repair Video Damaged by Network Data Loss," *Proceedings of IST/ACM/SPIE Multimedia Computing and Networking 2000 (MMCN00)*, Jan. 2000.
- [4] C. Perkins, I. Kouvelas, et al., "RTP Payload for Redundant Audio Data, IETF Audio/Video Transport Working Group," Sept. 1997. [RFC2198].
- [5] N. Feamster and H. Balakrishnan, "Packet Loss Recovery for Streaming Video," *12th International Packet VideoWorkshop*, Apr. 2002.
- [6] J. Klaue, 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*.
- [7] A. Lo, 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. 810, Jan. 2005.
- [8] R. W. Borko Furht, Joshua Greenberg, *Motion Estimation Algorithms For Video Compression*. Kluwer Academic Publishers, 1997.
- [9] A. B. Watson, J. Hu, and J. F. III. McGowan, "DVQ: A digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 2029, 2001.
- [10] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment. Synthesis Lectures on Image, Video and Multimedia Processing*. Morgan and Claypool, 2006.

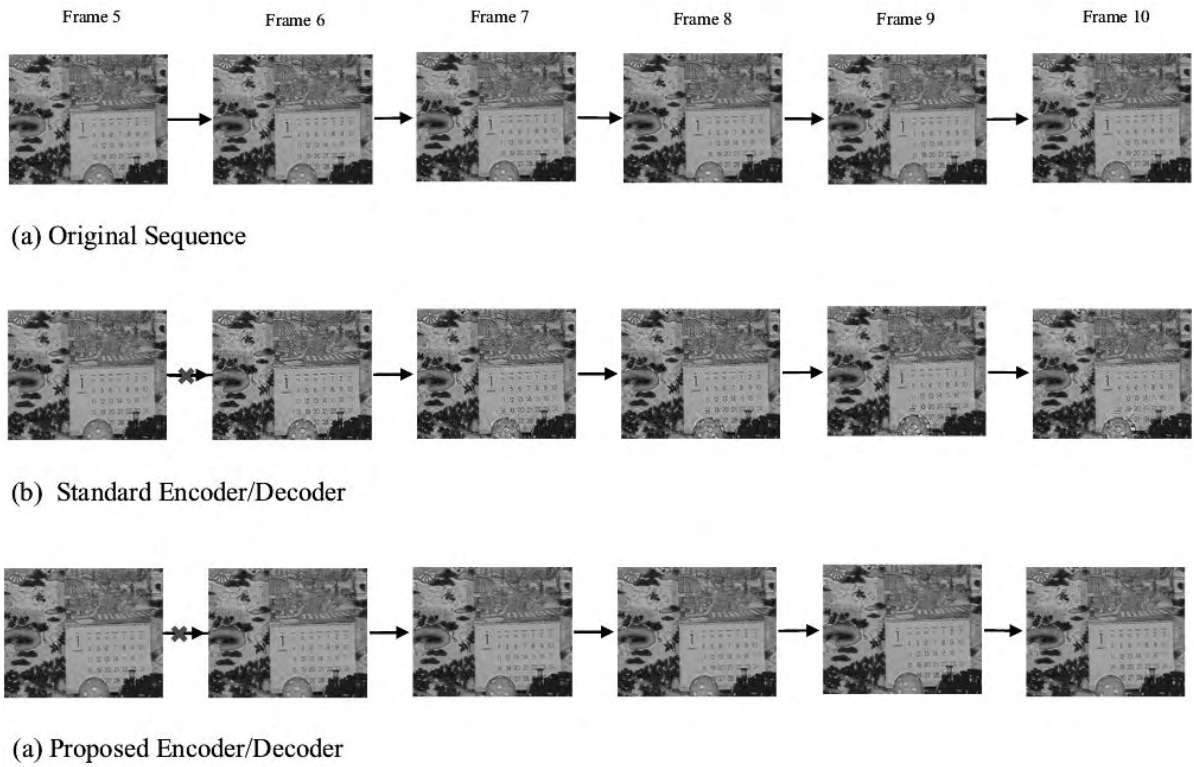


Fig. 4. Mobile QCIF sequence with the I-frame (frame 6) corrupted during transmission.

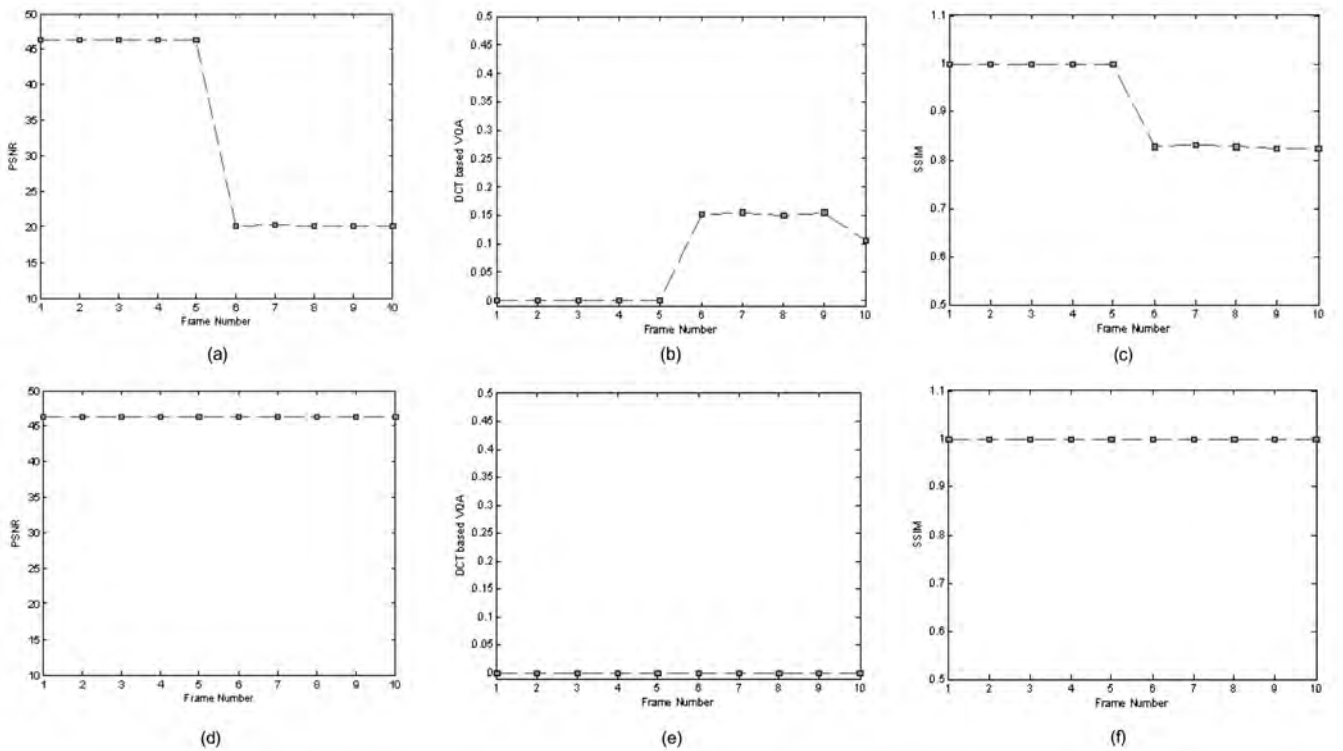


Fig. 5. Video Quality Analysis (using PSNR, DCT based metric and MSSIM) of the standard MPEG-4 decoder (a,b,c) and our proposed algorithm (d,e,f).

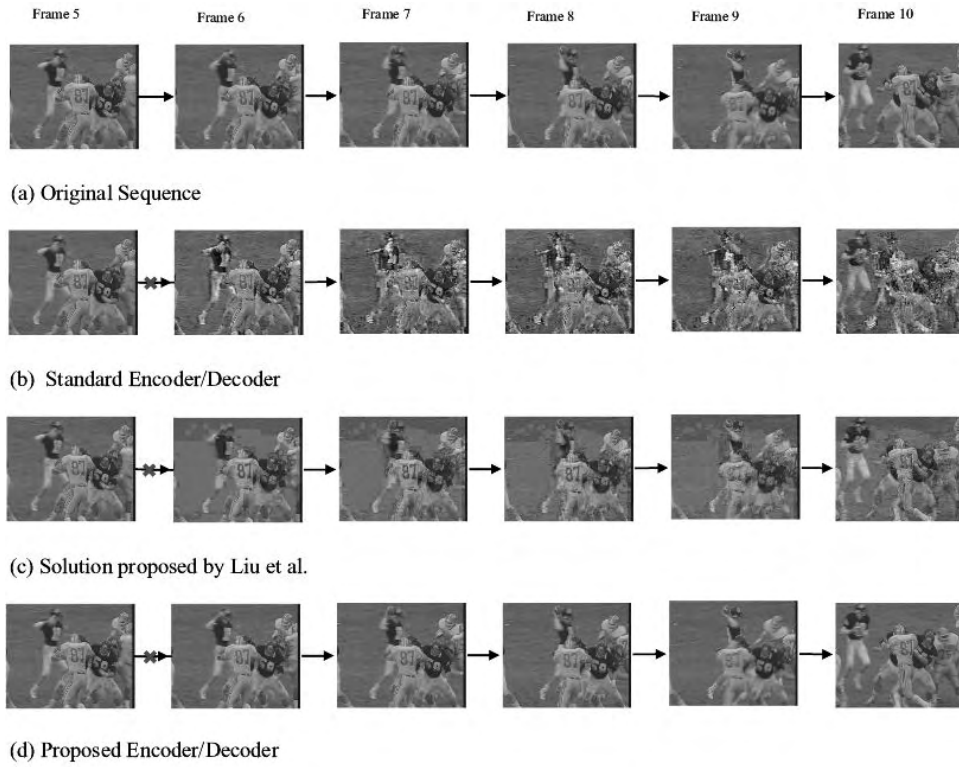


Fig. 6. Football CIF sequence with the I-frame (frame 6) corrupted during transmission.

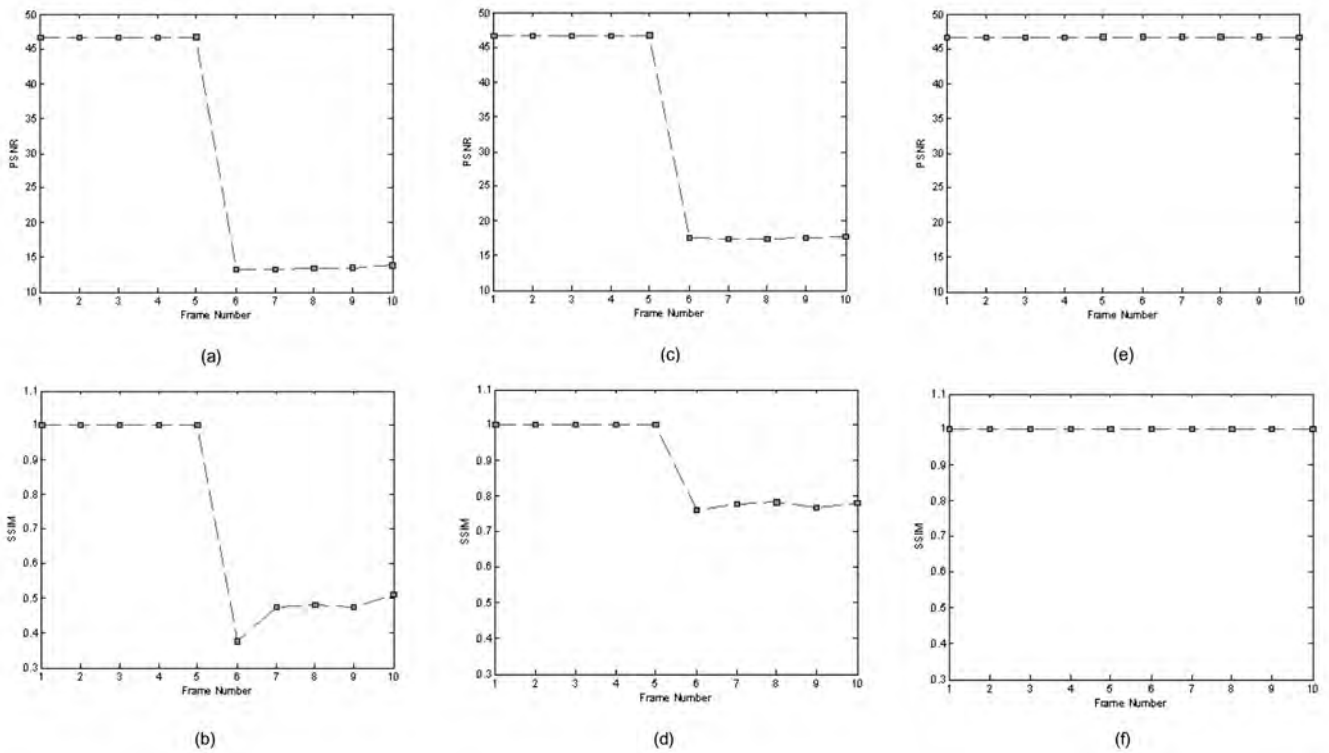


Fig. 7. Video Quality Analysis (using PSNR and MSSIM) of the standard MPEG-4 decoder (a,b), Solution proposed by [3] (c,d) and our proposed algorithm (e,f).