Motion Vector Recovery for Error Concealment based on Distortion Modeling

Jae-Won Suh and Yo-Sung Ho, *Member, IEEE*
Kwangju Institute of Science and Technology (K-JIST)
E-mail: won@gogh.kjist.ac.kr, hoyo@kjist.ac.kr

Abstract

If channel errors are introduced during real-time video transmission and cannot be corrected properly at the receiver, we can apply error concealment techniques to repair damaged portions of the frame by exploiting spatial and temporal redundancies in video signal. In general, compressed bitstreams generated by an MPEG video coder are quite sensitive to channel errors. If one packet is lost or received with errors, not only the current frame will be corrupted, but also the errors will propagate to succeeding frames. In this paper, we define a new model to describe the distortion of corrupted macroblocks in the bitstreams generated by an MPEG-2 video codec. Based on the proposed distortion model, we have developed several error concealment algorithms to reduce the effect of channel errors.

I. Introduction

Current video coding standards, such as H.261, H.263, MPEG-1, and MPEG-2, successfully achieve high compression ratios, employing discrete cosine transform (DCT), motion estimation (ME), motion compensation (MC), and variable length coding (VLC) techniques. Transmission of compressed video through various transmission channels may lead to information loss due to channel noises or traffic congestion. Compressed bitstreams generated by the MPEG-2 video coder are very sensitive to channel disturbances. Due to the nature of the MPEG-2 compressed video bitstream [1], a single bit error can affect not only the corresponding macroblock (MB), but also the whole frame, even further to the succeeding video frames. Therefore, if one or more coded bits are lost or corrupted during transmission, we need to estimate the lost data to obtain acceptable visual quality.

There are mainly two different approaches to reduce the transmission error effect. In order to detect and correct channel errors that occur during transmission, we can apply channel coding techniques, such as forward error correction (FEC) codes, in the transmission level. When we use mathematically well-structured codes, we can detect and even correct the transmission errors. However, this approach increases transmission bit rates due to added redundancies. Therefore, if we consider the limited channel bandwidth, FEC may not be always the best solution.

Alternatively, we can apply error concealment techniques to reconstruct the lost data by exploiting spatial and temporal redundancies existing in the video sequence. In the spatial-domain error concealment technique, we interpolate the lost information using adjacent luminance and color values around the erroneous MB [3,4]. In the temporal-domain error concealment technique, we estimate lost motion vector (MV) of the corrupted MB and compensate it by the estimated MV [5,6] under the assumption that MVs of spatially adjacent MBs are highly correlated.

In this paper, we define two MB distortion models for intra (I) and predictive (P) coded frames. The distortion model for the I-frame describes spatial dependencies in the current frame, and the distortion model for the P-frame exploits spatio-temporal dependencies between the current and the reference frames. Based on these models, we develop new temporal-domain error concealment techniques for MPEG-2 transport stream (TS) packets.

II. System Configuration

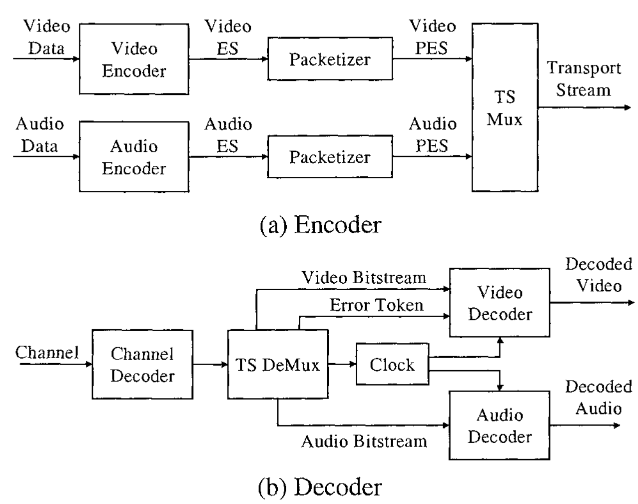


Figure 1. MPEG-2 TS System

Fig. 1 shows a block diagram of the general MPEG-2 system [2] for digital video communications. When MPEG-2 TS packets are transmitted over a noisy channel, some TS packets may be corrupted by various noises. Bit errors in the received bitstream can be detected at the decoder by several mechanisms. In our system, when the channel decoder detects any bit errors in the received bitstream, the TS DeMux sends an error token to the video decoder.

As shown in Fig. 2, the MPEG-2 TS packet consists of a 4-byte header and 184-byte payload data. The payload data include packetized elementary stream (PES) data. The header contains a 1-bit transport-error-indicator field, which indicates whether the received packet has one or more uncorrectable bit errors or not.

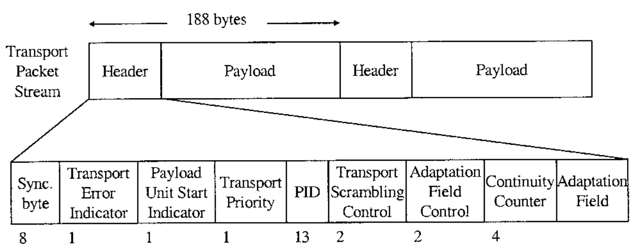
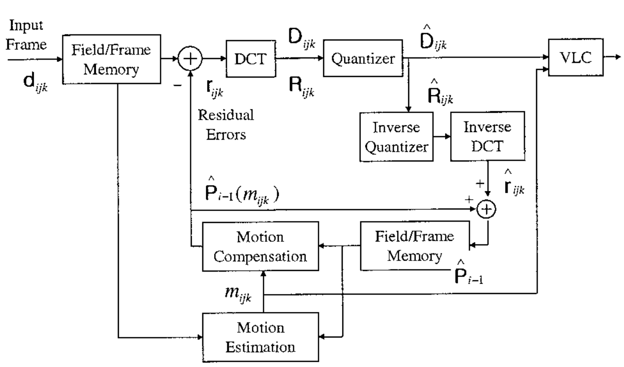
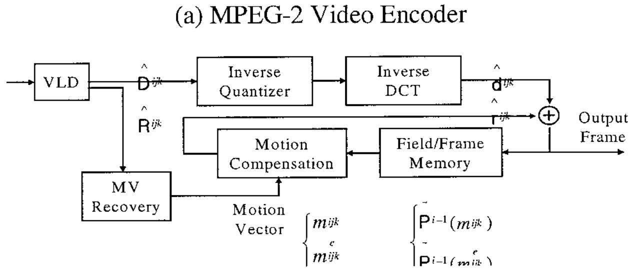


Figure 2. Transport Stream Syntax

If a received packet has uncorrectable bit errors, it is not easy to tell which bits are healthy in the packet. Therefore, for a practical purpose, the damaged packet is thrown out and treated as a lost packet. Furthermore, all successively received packets become useless until resynchronization is established. In MPEG-2, since the smallest resynchronization unit is the MB slice, we can lose horizontal strips of MBs from the first erroneous MB to the next slice start code. This coding structure helps to restrict error propagation to the MB slice within the image frame.



(a) MPEG-2 Video Encoder



(b) Modified MPEG-2 Video Decoder

Figure 3. Modified MPEG-2 Video Codec

Fig. 3(a) shows a block diagram of the MPEG-2 video encoder. For the I-frame coding, the image frame is partitioned into 8 × 8 blocks. After two-dimensional DCT is applied to each block, DCT coefficients are quantized. Then, the quantized DCT coefficients are coded using VLC. In the succeeding frames, temporal redundancies are removed by ME and MC. Residual errors and MV information are encoded and transmitted.

Fig. 3(b) shows a block diagram of a modified MPEG-2 video decoder. Except for MV recovery, the configuration is similar to the general MPEG-2 decoder. If we lose a MB, the MV recovery block estimates the MV of the lost MB and reconstruct for the lost MB using the estimated MV.

III. Distortion Model

In our distortion model, we consider a TS packet as a lost one if there is any error in the packet. Let S_{ij} be the event that the j-th slice of the i-th frame is corrupted. Then, we can define the slice error probability by

$$\Pr(S_{ij}) = \sum_{n=1}^{N_{ij}} {}_nC_k\, p_b^k (1-p_b)^{n-k} = 1-(1-p_b)^{N_{ij}} \qquad (1)$$

$$\text{where } i = 1, 2, \cdots, N,\ k \le n \text{ and } j = 1, 2, \cdots, J$$

where p_b is bit error probability, N_{ij} is the number of bits of the j-th slice of the i-th frame, J is the number of the slices of the frame, and N is the number of frames.

A. Distortion Model for the I-Frame

The I-frame has only intra-coded MBs. In Fig. 3(a) and Fig. 3(b), $\{ijk\}$ indicates the k-th MB in the j-th slice of the i-th frame. d_{ijk} is transformed into D_{ijk}. $\hat{\mathrm{D}}_{ijk}$ represents quantized MB data, and $\hat{\mathrm{d}}_{ijk}$ is correctly decoded MB data that include quantization distortion. As shown in Fig. 3(b), if we receive the corrupted MB data, concealed data $\tilde{\mathrm{d}}_{ijk}$ replaces $\hat{\mathrm{d}}_{ijk}$. Since temporal information is not available in the I-frame, we can estimate $\tilde{\mathrm{d}}_{ijk}$ only by spatial interpolation.

The distortion e_{ijk} of the I-frame can be represented by

$$\mathrm{e}_{ijk} = E[(\mathrm{d}_{ijk} - \hat{\mathrm{d}}_{ijk})^2](1-\Pr(S_{ij})) + E[(\mathrm{d}_{ijk} - \tilde{\mathrm{d}}_{ijk})^2]\Pr(S_{ij}) \qquad (2)$$

Since DCT is a unitary transform, the mean squared error (MSE) in the spatial domain is the same as that in the transform domain. The first term of Eq. (2) represents the quantization error. If we can assume that the quantization noise and the error concealment noise are uncorrelated, the second term of Eq. (2) can be decomposed into the quantization error and the pure reconstruction error by the error concealment technique. Therefore,

$$E[(\mathrm{d}_{ijk} - \tilde{\mathrm{d}}_{ijk})^2] = E[(\mathrm{d}_{ijk} - \hat{\mathrm{d}}_{ijk})^2] + E[(\hat{\mathrm{d}}_{ijk} - \tilde{\mathrm{d}}_{ijk})^2] \qquad (3)$$

By Eq. (3), we can rewrite Eq. (2) as

$$e_{ijk} \approx E[(d_{ijk} - \hat{d}_{ijk})^2] + E[(\hat{d}_{ijk} - \tilde{d}_{ijk})^2]\Pr(S_{ij}) \qquad (4)$$

We can employ different error concealment algorithms to obtain $\tilde{d}_{ijk}$ from adjacent MBs around the lost MB.

B. Distortion Model for the P-Frame

In this section, we define MB distortion of the first P-frame in a group of pictures (GOP). The P-frame has both intra- and inter-coded MBs. The superscript I or P denotes intra- or inter-coded mode, respectively.

r_{ijk} denotes motion-compensated residual errors, and R_{ijk} is DCT transformed data of the residual errors. $\hat{R}_{ijk}$ is quantized data of R_{ijk}, and m_{ijk} represents a MV. $\hat{P}_{i-1}$ and $\tilde{P}_{i-1}$ indicate reference frames that are used for ME and MC at the encoder and the decoder, respectively, where i is the coding order. They are different because some parts of the reference I-frame at the decoder might be concealed. $\tilde{P}_{i-1}(m_{ijk})$ is a motion-compensated MB by MV m_{ijk}. When we receive corrupted data, we estimate the lost MV from $\tilde{P}_{i-1}$ and compensate for the lost MB using the estimated MV m^e_{ijk}.

The distortion e^P_{ijk} of the P-frame can be expressed as

$$\begin{aligned} e^P_{ijk} &= E[\{(r_{ijk} + \hat{P}_{i-1}(m_{ijk})) - \tilde{P}_{i-1}(m^e_{ijk})\}^2]\Pr(S_{ij}) \\ &+ E[\{(r_{ijk} + \hat{P}_{i-1}(m_{ijk})) - (\hat{r}_{ijk} + \tilde{P}_{i-1}(m_{ijk}))\}^2](1 - \Pr(S_{ij})) \end{aligned} \qquad (5)$$

If we receive corrupted MB data, we cannot use the residual errors to reconstruct current MB in the first term of Eq. (5). In order to separate meaningful terms, if we assume that the quantization noise of the residual errors and the error concealment noise are uncorrelated, the first term of Eq. (5) can be rearranged as

$$\begin{aligned} &E[(\hat{r}_{ijk} + \hat{P}_{i-1}(m_{ijk}) - \tilde{P}_{i-1}(m^e_{ijk}))^2]\Pr(S_{ij}) \\ &\quad + E[(r_{ijk} - \hat{r}_{ijk})^2]\Pr(S_{ij}) \end{aligned} \qquad (6)$$

If we use another assumption that the quantization noise of the residual errors and the mismatch between reference frames are uncorrelated, the second term of Eq. (5) can be represented as

$$\begin{aligned} &E[(\hat{P}_{i-1}(m_{ijk}) - \tilde{P}_{i-1}(m_{ijk}))^2](1 - \Pr(S_{ij})) \\ &\quad + E[(r_{ijk} - \hat{r}_{ijk})^2](1 - \Pr(S_{ij})) \end{aligned} \qquad (7)$$

By Eq. (6) and Eq. (7), we can rewrite Eq. (5) as

$$\begin{aligned} e^P_{ijk} &\approx E[\{r_{ijk} - \hat{r}_{ijk}\}^2] \\ &+ E[\{\hat{P}_{i-1}(m_{ijk}) - \tilde{P}_{i-1}(m_{ijk})\}^2](1 - \Pr(S_{ij})) \\ &+ E[\{\hat{r}_{ijk} + \hat{P}_{i-1}(m_{ijk}) - \tilde{P}_{i-1}(m^e_{ijk})\}^2]\Pr(S_{ij}) \end{aligned} \qquad (8)$$

The distortion e^I_{ijk} of the P-frame can be represented as

$$\begin{aligned} e^I_{ijk} &= E[(d_{ijk} - \hat{d}_{ijk})^2](1 - \Pr(S_{ij})) \\ &+ E[(d_{ijk} - \tilde{d}_{ijk})^2]\Pr(S_{ij}) \\ &\approx E[(d_{ijk} - \hat{d}_{ijk})^2](1 - \Pr(S_{ij})) \\ &+ E[\{d_{ijk} - \tilde{P}_{i-1}(m^e_{ijk})\}^2]\Pr(S_{ij}) \end{aligned} \qquad (9)$$

We can obtain $\tilde{d}_{ijk}$ from the reference frame and the estimated MV value using temporal redundancies. From Eq. (8) and Eq. (9), if we estimate more accurate MV, we can obtain good performance.

IV. New Error Concealment Schemes

When we have bit errors in the received bitstream, we have to throw out the erroneous packet and the succeeding packets until we detect the next MB slice header. Therefore, we may lose all the MBs in the slice and have only the upper and lower MBs of the lost MB slice. We also note that some MBs are intra-coded; therefore, they do not have MVs.

The MV of the lost MB can be estimated by taking the average value of MVs of vertically adjacent MBs [5]. If the upper and the lower MBs are inter-coded, the MV of the lost MB is the average of the two MVs of the neighboring MBs. If only one of the vertical neighbors has a valid MV, we use the MV for the lost MB. Unfortunately, if none of the vertical neighbors has a valid MV, we assume that the MV is zero. In this scheme, if vertically neighboring MBs have MVs, we can obtain reasonably good reconstruction quality for the lost MB. However, if only one or none of the vertical neighbors has a valid MV, quality of the reconstructed image is not satisfactory.

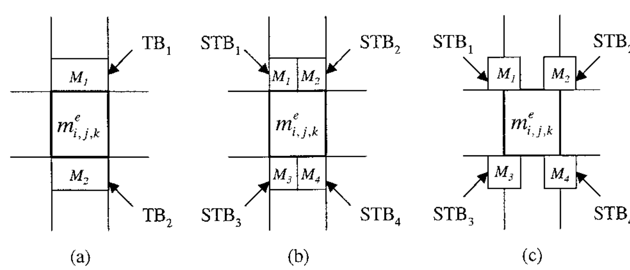
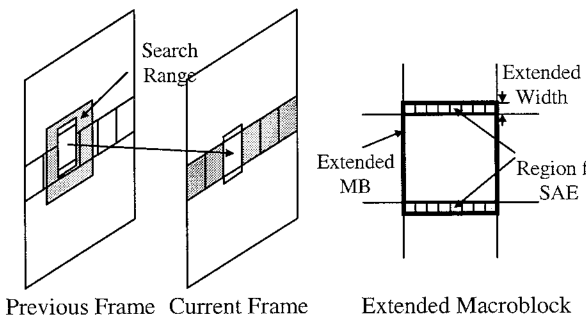


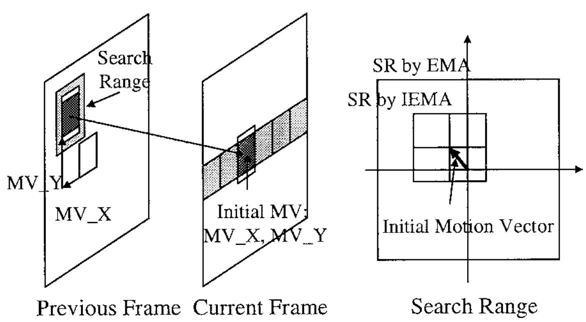
Figure 4. Modified Average Algorithm

In order to solve the above problem, we devise the following schemes for recovering the lost MBs. As shown in Fig. 4(a), we define 16 x 8 target blocks (TBs) in adjacent MBs. In order to obtain an accurate MV for the lost MB, we need that vertically adjacent MBs have MVs. For example, if only one MB between vertically neighboring MBs has a MV, the MV of the other TB is estimated by a block matching algorithm (BMA) at the decoder. Then, we can recover the lost MV by taking average of the estimated MV of the adjacent TB and the original MV of the adjacent MB.

In order to obtain a more accurate MV for the lost MB, we separate 16 x 8 TB into two 8 x 8 small target blocks (STBs) and take average of estimated MVs for STBs and the original MV of the adjacent MB, as shown in Fig. 4(b). These two algorithms, however, may increase computational complexity at the decoder considerably due to the motion estimation operation. In order to reduce the computation burden at the decoder, we define alternative STBs, as shown in Fig. 4(c). With this modification, we can reduce a half of the computational complexity, compared to Fig. 4(b).



(a) Extension Matching Algorithm (EMA)



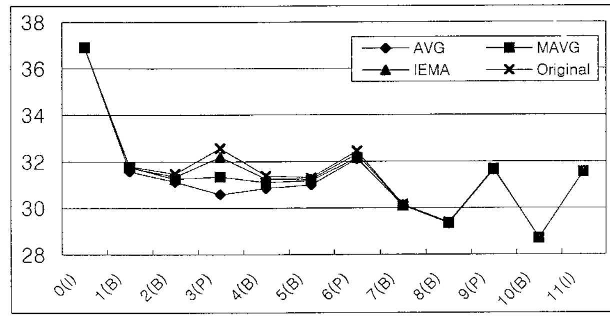
(b) EMA with Initial MV (IEMA)

Figure 5. Extension Matching Algorithm

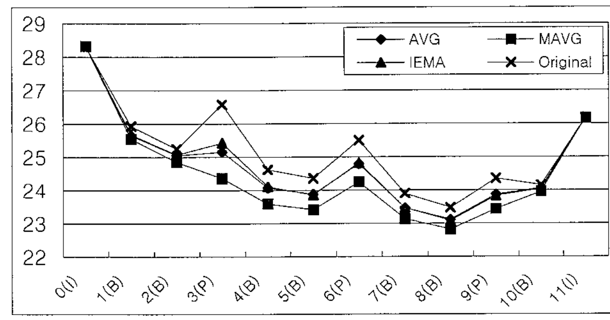
We can also estimate the MV of the lost MB by noting that neighboring pixels often move together. As shown in Fig. 5(a), we form an extended MB. The lost MV can be determined by minimizing the sum of absolute errors (SAE) between extended pixels in the current frame and the previous reference frame. In this algorithm, the extended width and the search range are very important parameters. If the extended width and the search range are increased, so is the computational time required. Therefore, EMA entails a considerable amount of processing complexity at the decoder.

In order to reduce the computational time of EMA, we can use an initial MV. We set the initial MV for the lost MB using the average value of MVs of the vertically adjacent MBs. The initial MV establishes a starting point of the motion search. By setting the initial MB, we can reduce the search range and the processing time requirements effectively, as shown in Fig. 5(b).

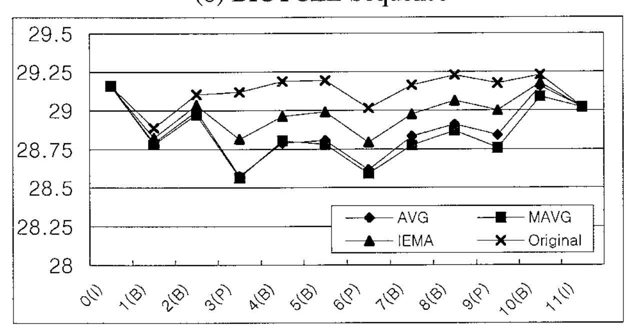
V. Simulation Results



(a) FOOTBALL Sequence



(b) BICYCLE Sequence



(c) BALLET Sequence

Figure 6. Performance Comparison

In order to evaluate the performance of various error concealment algorithms, we perform computer simulations on several test sequences: FOOTBALL, BICYCLE, and BALLET. They have the 4:2:0 format of 720 x 480 pixels. Each GOP consists of 12 frames (N=12, M=3).

Fig. 6 plots average PSNR values of reconstructed frames for the test sequences. We assume that the first P-frame has lost one packet. In Fig. 6, the x-axis indicates the frame number, and the y-axis represents the PSNR values. We have compared performance of four algorithms: an average algorithm (AVG) by [5], a modified average algorithm (MAVG) that has [-25, 25] search range and 8 x 8 STB size, an extension matching algorithm (EMA) where the extended width is 1 and the search range is [–25, 25], and an extension matching algorithm with an initial MV (IEMA) where the extended width is 1 and the search range is reduced to [–5, 5].

Fig. 7 shows a corrupted P-frame of the FOOTBALL sequence and concealed frames by MC with estimated MV using various MV recovery algorithms. Fig. 7(b) is obtained by AVG. It gives good concealment performance in small and slow motion areas, but shows some blocky spots in large and fast motion areas. Fig. 7(c) and Fig. 7(d) are obtained by MAVG and IEMA. From the PSNR results and the subjective quality comparison, we have observed that IEMA shows relatively better performance.

VI. Conclusions

In this paper, we have defined a distortion model for error concealment and proposed various error concealment algorithms based on the distortion model. Temporal-domain error concealment algorithms provide well-concealed data by replacing the lost data by the data in the previous reference frame. Although the average algorithm provides good reconstruction in stationary regions, we can find degradations in fast moving areas. In order to solve this problem, we have proposed a modified average algorithm, an extension matching algorithm, and an extension matching algorithm with an initial MV. Although the modified average algorithm demonstrates substantial improvement in PSNR values and reconstruction quality, it requires high computational complexity. In order to reduce the processing time, we have proposed an extension matching algorithm with an initial MV, which has shown improve reconstruction quality and high PSNR values.

Acknowledgement

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) through the Ultra-Fast Fiber-Optic Networks (UFON) Research Center at Kwangju Institute of Science and Technology (K-JIST), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

[1] ISO/IEC IS 13818-2 (MPEG-2 Video): Information Technology - Generic Coding of Moving Pictures and Associated Audio Information, April 1996.

[2] ISO/IEC IS 13818-1 (MPEG-2 System): Information Technology- Generic Coding of Moving Pictures and Associated Audio Information, April 1996.

[3] S. Aign and K. Fazel, "Temporal and Spatial Error Concealment Techniques for Hierarchical MPEG-2 Video Codec," *IEEE International Conference on Communication*, Vol. 3, pp. 1778-1783, 1995.

[4] J.W. Suh and Y.S. Ho, "Error Concealment Based on Directional Interpolation," *IEEE Trans. on Consumer Electronics*, Vol. 43, No. 3, pp. 295-302, Aug. 1997.

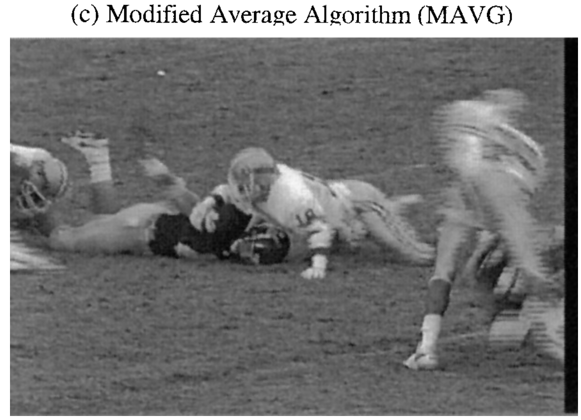
[5] H. Sun, K. Challapali and J. Zdepski, "Error Concealment in Digital Simulcast AD-HDTV Decoder," *IEEE Trans. on Consumer Electronics*, Vol. 38, No. 3, pp. 108-116, Aug. 1992.

[6] J. Feng, K. Lo, H. Mehropour and A.E. Karbowiak, "Error Concealment for MPEG Video Transmissions," *IEEE Trans. on Consumer Electronics*, Vol. 43, pp. 183-187, May 1997.

(a) Corrupted P-Frame

(b) Average Algorithm (AVG)

(c) Modified Average Algorithm (MAVG)



(d) EMA with Initial MV (IEMA)

Figure 7. Subjective Quality Evaluation