Error Concealment in Encoded Video Streams*

Paul Salama, Ness B. Shroff, Edward J. Delp
Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Corresponding Author:
Professor Edward J. Delp
School of Electrical and Computer Engineering
1285 Electrical Engineering Building
Purdue University
West Lafayette, IN 47907-1285
USA
Telephone: +1 765 494 1740
Fax: +1 765 494 0880
Email: ace@ecn.purdue.edu

Abstract

When transmitting compressed video over a data network, one has to deal with how channel errors affect the decoding process. This is particularly problematic with data loss or erasures. In this paper we describe techniques to address this problem in the context of networks where channel errors or congestion can result in the loss of entire macroblocks when MPEG video is transmitted. We describe spatial and temporal techniques for the recovery of lost macroblocks. In particular, we develop estimation techniques for the reconstruction of missing macroblocks using a Markov Random Field model. We show that the widely used heuristic motion compensated error concealment technique based on averaging motion vectors is a special case of our estimation technique. We further describe a technique that can be implemented in real-time.

Index Terms: ATM, cell loss, error concealment, motion vectors, Markov random field, cell packing, spatial reconstruction, temporal reconstruction.

---

1. INTRODUCTION

When transmitting compressed video over a data network, one has to deal with how channel errors affect the decoding process. This is particularly a problem with data loss or erasures. In this paper we describe techniques to address this problem.

It is envisioned that one of the most important network applications will involve transmitting digital video [1]. During periods of network congestion packets may be dropped badly degrading the quality of the video, as a result of the missing data. Since retransmission is not a viable option for real-time multimedia applications, error concealment algorithms need to be developed. These algorithms estimate the missing data from the received video in an effort to conceal the effect of channel impairements. Furthermore, these algorithms will have to be integrated into the decoder hardware, and hence be simple enough to be implemented in real time.

In this paper we derive error concealment techniques based on a Markov Random Field (MRF) model [2, 3, 4]. The algorithms are categorized as being either spatial (utilizing pixel data) or temporal (utilizing motion vectors) in nature. The spatial techniques rely on pixel data within a current damaged frame and an MRF model of the frame to restore damaged areas of the frame. Similarly, estimates for missing motion vectors are obtained by modeling the motion field as a MRF and finding the maximum a posteriori (MAP) estimate of each missing motion vector given its neighboring motion vectors. We also show that the widely used huerisitic technique based on averaging the motion vectors of neighboring macroblocks [5] is a special case of our MAP estimate. We also describe a Temporal-Spatial method for restoring damaged macroblocks based on the use of a ternary tree for classifying the available motion vectors and the MRF model for the corrupted frame.

The paper is organized as follows: the concealment of lost macroblocks resulting from ATM cell loss is addressed in Section 3. The performance and comparison of our error concealment algorithms are provided in Section 4, and the conclusion of our work is given in Section 5.

2.   Previous Work

Numerous schemes have been proposed in the literature to combat the effects of data loss on encoded video. Currently the approaches utilized for signal restoration/error concealment are either active concealment or passive concealment. In active concealment, error control coding techniques are used along with retransmission. Since extra data must be transmitted, it is sometimes necessary to reduce the source coder's data rate to avoid increasing network congestion. Active concealment has the advantage of permitting perfect reconstruction at the decoding end, if the amount of data lost is not significant, i.e. within the parameters of the error control coding scheme. In addition, unequal error protection can be provided by varying the number of bits used according to the priority of the data being protected [6, 7]. In passive concealment the video stream is post-processed to reconstruct the missing data. Although passive concealment does not result in perfect reconstruction of the lost data, it is necessary in many applications where error control coding cannot be used because of a high level of overhead, problem with compliance with video transmission standards, or when active concealment itself fails. A general overview with an extensive bibliography of the various error concealment methods can be found in [8].

Since video sequences that have been compressed according to the current video compression standards [9-12] consist of DCT coefficients and motion vectors, one approach to error concealment over networks is to prioritize data as described in [13-20]. Encoded video data is segmented into low priority data such as high frequency DCT coefficients, and high priority data such as addresses of blocks, motion vectors, and low frequency DCT coefficients. In the event of network congestion, packets carrying low priority data are discarded, while those carrying high priority data are retained.

Alternative techniques for reducing the effect of packet loss were proposed in [16, 22]. In [16] the data rate of the encoded bitstream is decreased when the packet loss rate increases. A similar approach for subband coded sequences is described in [22], wherein error correcting codes are employed to correct data lost due to packet loss. Since the use of error correcting codes increases the data rate, the source coder is throttled such that the ultimate data rate is equal to the data rate of the original unprotected video sequence. If the packet loss rate is

too severe and the lost data cannot be retrieved by the error correcting code, the damaged frame areas are then filled in with data from the previous frame.

Since packet loss can result in the loss of entire rows of macroblocks in an image, packetization techniques that rely on interleaving data have been proposed in [17, 23, 24, 25]. Interleaving is performed either at the macroblock level [23, 24], or at the Slice level [17]. This has the advantage of distributing the error due to packet loss over the entire frame rather than localizing it, and hence aids passive concealment techniques that rely entirely on spatial data to restore the damaged regions. It however, increases decoder complexity.

Post processing (passive) techniques for the sake of error concealment utilize spatial data, or temporal data, or a hybrid of both [5,24,26-39]. Missing macroblocks can be reconstructed by estimating their low frequency DCT coefficients from the DCT coefficients of the neighboring macroblocks [24, 30, 31, 32], by estimating missing edges in each block from edges in the surrounding blocks as proposed in [34], or by the method of projections onto convex sets [40] as described in [35].

An alternative to using spatial data for error concealment is to use motion compensated concealment [5] whereby the average of the motion vectors of neighboring macroblocks is used to perform concealment. We will show later that this is a special case of one of our motion compensated techniques.

Currently the periodic insertion of Resynchronization Markers (unique codes that limit the effect of data loss and establish synchronization between encoder and decoder) is being considered for the MPEG-4 and JPEG2000 standards [41, 42]. The resilience of MPEG-4 bitstreams is also increased by placing coded motion vector data prior to coded DCT coefficients. These two groups of data are separated by a field known as the Motion Boundary Marker (MBM) [41, 42]. The MBM is utilized by the decoder to discern whether coded motion vectors have been corrupted or not. Uncorrupted coded motion vectors are used to reconstruct macroblocks. If however the motion vector data between two resynchronization markers has been corrupted, all the data between the two markers is discarded and the macroblocks constructed via temporal replacement from the previous frame. The robustness of the stream is further enhanced by using reversible variable length codes that can be uniquely forward or backward decoded, as well as by repeating important header information [41, 42].

3.   PASSIVE ERROR CONCEALMENT

The goal of passive error concealment is to estimate missing data. In the case of MPEG video, the objective is to estimate missing macroblocks and motion vectors. The underlying idea is that there is still enough redundancy in the sequence to be exploited by the concealment technique. In particular, in I frames it is possible to have a lost macroblock surrounded by intact macroblocks that are used to interpolate the missing data. This is a result of the fact that macroblocks in I frames can span across two packets. It also arises when the macroblocks in I frames are interleaved prior to packing them into packets. In P and B pictures, it is possible to have entire rows of macroblocks missing. In this case, spatial interpolation will not yield acceptable reconstructions. However, the motion vectors of the surrounding regions can be used to estimate the lost vectors, and the damaged region reconstructed via motion compensated interpolation [27, 28].

Let X be an $N_1 \times N_2$ decompressed frame from an MPEG sequence and let Y be the received version at the output of the channel. Due to channel errors it is conceivable that Y may have missing data. Each transmitted picture consists of M macroblocks that have $N \times N$ pixels. Let $\mathbf{x}_i$ be the lexicographic ordering of the pixels in the ith macroblock in X. The vector $\mathbf{x}$ is then defined to be the concatenation of $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_M$, that is, $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots \ \mathbf{x}_M^T]^T$. The vector $\mathbf{y}$ is similarly defined for Y. If the jth macroblock is missing due to packet loss then $\mathbf{y} = \mathbf{D}\mathbf{x}$, where $\mathbf{D}$ is an $(N_1 N_2 - N^2) \times N_1 N_2$ matrix that consists of the identity matrix excluding the rows from row jN^2 to row $(j+1)N^2 - 1$. If n of the M macroblocks are missing due to packet loss, then $\mathbf{D}$ will be an $(N_1 N_2 - nN^2) \times N_1 N_2$ matrix. The goal of passive error concealment is to estimate $\mathbf{x}$ given the received data $\mathbf{y}$. A description of how this is done will be given in the following sections.

3.1.   Statistical Spatial Approach: MAP Estimation

Statistical techniques have been successfully used in image processing for edge reconstruction [4, 43, 44, 45, 46]. The original image is modeled as a Markov random field (MRF) [2, 3, 4], and edges are reconstructed by maximum *a posteriori* (MAP) techniques. This is the approach adopted here. Each original frame X and its received version Y are modeled as discrete parameter random fields where each pixel is a continuous random variable. Assum-

ing a prior distribution for $\mathbf{x}$, a maximum *a posteriori* (MAP) estimate is obtained given the received data $\mathbf{y}$. Denoting the estimate of $\mathbf{x}$ by $\hat{\mathbf{x}}$, $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}|\mathbf{y}=\mathbf{Dx}} L(\mathbf{x} \mid \mathbf{y})$, where $L(\mathbf{x} \mid \mathbf{y})$ is the log-likelihood function. In other words, $L(\mathbf{x} \mid \mathbf{y}) = \log f(\mathbf{x} \mid \mathbf{y})$, where $f(\mathbf{x} \mid \mathbf{y})$ is the conditional probability density function of $\mathbf{x}$ given $\mathbf{y}$. It is shown in Appendix A that

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}|\mathbf{y}=\mathbf{Dx}} [-\log f(\mathbf{x})]. \tag{1}$$

Since X is modeled as a Markov Random Field (MRF), the probability density function of $\mathbf{x}$ is given by [2, 4, 47],

$$f(\mathbf{x}) = \frac{1}{Z} \exp\left(-\sum_{c\epsilon C} V_c(\mathbf{x})\right), \tag{2}$$

where Z is a normalizing constant known as the *partition function*, $V_c(\cdot)$ a function of a local group of points c known as cliques, and C the set of all cliques [4]. Using Equations (1) and (2), the MAP estimate of $\mathbf{x}$ is then

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}|\mathbf{y}=\mathbf{Dx}} \left[\sum_{c\epsilon C} V_c(\mathbf{x})\right].$$

In the following we will indicate our choice of the potential functions [4] and describe the process by which we obtain the MAP estimate.

3.1.1. THE MAP ESTIMATE

The proper choice of V_c is crucial in the reconstruction of the macroblocks. In this case the potential functions are chosen such that

$$\sum_{c\epsilon C} V_c(\mathbf{x}) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \sum_{m=0}^{3} b_{i,j}^{(m)} \rho(\frac{D_m(X_{i,j})}{\sigma}), \tag{3}$$

where $D_0(X_{i,j}) = X_{i,j-1} - X_{i,j}$, $D_1(X_{i,j}) = X_{i-1,j+1} - X_{i,j}$, $D_2(X_{i,j}) = X_{i-1,j} - X_{i,j}$, and $D_3(X_{i,j}) = X_{i-1,j-1} - X_{i,j}$ are used to approximate the first order derivatives at pixel i, j. $\rho(\cdot)$ is a cost function, σ a scaling factor, $b_{i,j}^{(m)}$ weighting coefficients, and the set of cliques [4] is $C = \{ \{(i, j-1), (i, j)\}, \{(i-1, j+1), (i, j)\}, \{(i-1, j), (i, j)\}, \{(i-1, j-1), (i, j)\} \}$.

Several cost functions have been proposed [43, 45]. A convex $\rho(\cdot)$ results in the minimization of a convex functional. The cost function used here is the one introduced by Huber for obtaining robust M-estimates of location [48]. Its advantage is that it is convex, does not

heavily penalize edges, and is simpler to implement than most of the convex cost functions used in the literature [45]. It is defined to be

$$\rho_\gamma(x) = \begin{cases} x^2 & |x| \le \gamma \\ \gamma^2 + 2\gamma(|x| - \gamma) & |x| > \gamma. \end{cases}$$

Hence, $\sum_{c \epsilon C} V_c(\mathbf{x}) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \sum_{m=0}^{3} b_{i,j}^m \rho_\gamma(\frac{D_m(X_{i,j})}{\sigma})$. Letting

$$h_\gamma(\mathbf{x}) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} \sum_{m=0}^{3} b_{i,j}^m \rho_\gamma(\frac{D_m(X_{i,j})}{\sigma}),$$

we have

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}|\mathbf{y}=\mathbf{Dx}} h_\gamma(\mathbf{x}). \tag{4}$$

The solution to Equation (4) above can be obtained by means of the iterative conditional modes (ICM) algorithm [49]. In particular if the ith element of $\mathbf{x}$, denoted by x_i, corresponds to a lost pixel value, and $\mathbf{x}_{\partial i}$ denotes the neighborhood of x_i then

$$\hat{x}_i = \arg \max_{x_i} f(x_i \mid \mathbf{x}_{\partial i}). \tag{5}$$

Using Equations (3) and (5), the MAP estimate of pixel (i, j), given its neighbors, is

$$\hat{X}_{i,j} = \arg \min_{X_{i,j}} \sum_{l=i}^{i+1} \sum_{k=j}^{j+1} \sum_{m=0}^{3} b_{l,k}^m \rho_\gamma(\frac{D_m(X_{l,k})}{\sigma}). \tag{6}$$

Let $\mathbf{X}_{\partial i}$ denote the neighborhood of the ith macroblock $\mathbf{x_i}$. The MAP estimate of $\mathbf{x_i}$ satisfies $\hat{\mathbf{x}}_\mathbf{i} = \arg \max_\mathbf{x} f(\mathbf{x_i} \mid \mathbf{X}_{\partial i})$. If we let $\mathbf{J_i}$ denote the set of indices of the pixels belonging to $\mathbf{x_i}$, then it can be similarly shown that

$$\hat{\mathbf{x}}_\mathbf{i} = \arg \min_\mathbf{x_i} \sum_{(i,j)\epsilon\mathbf{J_i}} \sum_{l=i}^{i+1} \sum_{k=j}^{j+1} \sum_{m=0}^{3} b_{l,k}^m \rho_\gamma(\frac{D_m(X_{l,k})}{\sigma}). \tag{7}$$

The solution to Equation (7) can be obtained iteratively. This however is computationally intensive. In the subsequent section we describe how we speed up the process of finding a solution by using median filtering techniques to obtain a suboptimal MAP estimate.

3.2. Median Filtering: A Suboptimal Approach

The choice of γ and σ is crucial to the reconstruction of edges. The smaller the product $\gamma\sigma$, the less the edges are penalized. Since $h_\gamma(\mathbf{x})$ is continuous, convex, and has continuous first partial derivatives, then by successively iterating with respect to each pixel, a global minimum is attained.[1] Using Equation (6) we obtain,

$$\frac{\partial}{\partial X_{i,j}}h_\gamma(\mathbf{x}) = \frac{b_{i,j+1}^0}{\sigma}\rho_\gamma'(\frac{D_0(X_{i,j+1})}{\sigma}) + \frac{b_{i+1,j-1}^1}{\sigma}\rho_\gamma'(\frac{D_1(X_{i+1,j-1})}{\sigma})$$
$$+ \frac{b_{i+1,j}^2}{\sigma}\rho_\gamma'(\frac{D_2(X_{i+1,j})}{\sigma}) + \frac{b_{i+1,j+1}^3}{\sigma}\rho_\gamma'(\frac{D_3(X_{i+1,j+1})}{\sigma})$$
$$- \sum_{m=0}^{3}\frac{b_{i,j}^m}{\sigma}\rho_\gamma'(\frac{D_m(X_{i,j})}{\sigma}),$$

where

$$\rho_\gamma'(x) = \begin{cases} 2x & |x| \leq \gamma \\ 2\gamma & x > \gamma \\ -2\gamma & x < -\gamma. \end{cases} \tag{8}$$

Each pixel in the interior, has 8 neighbors. Let $z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8$ be the 8 neighbors arranged in ascending order, and rename the associated weights $\{b_{l,k}^m \mid l = i, i+1,\ l = j, j+1,\ m = 0\cdots3\}$ as $b_1, b_2, b_3,\ b_4, b_5, b_6, b_7, b_8$. Defining

$$U = \{(k,l) \mid k = i-1 \text{ and } l = j-1, j, j+1 \text{ or } k = i \text{ and } l = j-1\},$$
$$L = \{(k,l) \mid k = i+1 \text{ and } l = j-1, j, j+1 \text{ or } k = i \text{ and } l = j+1\},$$

and

$$\Delta_k(X_{i,j}) = \begin{cases} z_k - X_{i,j} & z_k \epsilon U \\ X_{i,j} - z_k & z_k \epsilon L, \end{cases}$$

then

$$\frac{\partial}{\partial X_{i,j}}h_\gamma(\mathbf{x}) = \frac{1}{\sigma}\sum_{k\epsilon L}b_k\rho_\gamma'(\frac{\Delta_k(X_{i,j})}{\sigma}) - \frac{1}{\sigma}\sum_{k\epsilon U}b_k\rho_\gamma'(\frac{\Delta_k(X_{i,j})}{\sigma}). \tag{9}$$

Since, we are iterating for $X_{i,j}$, we need to solve

$$\frac{\partial}{\partial X_{i,j}}h_\gamma(\mathbf{x}) = 0, \tag{10}$$

---

[1]There is, however, more than one global minimum.

When solving Equation (10), three cases need to be considered.

Case1: $|\mathbf{\Delta_k}| \leq \gamma\sigma, \ \forall \mathbf{k}$

This occurs when $z_8 - \gamma\sigma \leq z_1 + \gamma\sigma$, and the optimum value of $X_{i,j}$ satisfies the constraint: $|\Delta_k| \leq \gamma\sigma \ \forall k$. Hence, $h_\gamma(\hat{X}_{i,j}) = \sum_{k=1}^{8} b_k (\frac{\Delta_k}{\sigma})^2$. Using Equations (10) and (8)

$$\hat{X}_{i,j} = \frac{\sum_{k=1}^{8} b_k z_k}{\sum_{k=1}^{8} b_k}. \tag{11}$$

Case2: $|\mathbf{\Delta_k}| \leq \gamma\sigma$ for some k

We show in Appendix B, that the optimum estimate $\hat{X}_{i,j}$ satisfies

$$\hat{X}_{i,j} = \frac{\sum_{k=J_1}^{J_2} b_k z_k + \gamma\sigma[\sum_{k=J_2+1}^{8} b_k - \sum_{k=1}^{J_1-1} b_k]}{\sum_{k=J_1}^{J_2} b_k}$$

where J_1 and J_2 satisfy $z_{J_2} - z_{J_1} \leq \gamma\sigma, \ J_2 \geq J_1$.

We fix σ and choose γ to be arbitrarily small and positive. Then according to **Case 2A** in Appendix B, if equal weights $\{b_{l,k}^m \mid m = 0 \cdots 3, \ l = i, i+1, \ k = j, j+1\}$ are used, one possible choice for the optimum value $\hat{X}_{i,j}$ will be the median of its neighbors, unless there are at least two neighboring pixels that are equal in value. Under such conditions, the common pixel value is used. Although this is a suboptimal strategy, the resulting reconstruction technique is faster than searching for the optimum value. The estimate of each missing pixel value is now obtained by finding the median of 8 values instead of performing line search techniques.

3.3.   Estimation of Boundary Pixels

The above mentioned techniques coupled with macroblock or Slice [9] interleaving are particularly useful in the restoration of intracoded macroblocks that belong to a frame that serves as the anchor frame for a new scene within the same sequence.

An alternative approach that is useful for reconstructing intra coded macroblocks when the current damaged frame and the previous reference frame belong to the same scene is next described. The underlying idea is to try to find a macroblock sized region in the previous frame, X^{-1}, that will maximize the MAP estimate of the boundary pixels of the missing macroblock given its neighbors.

Formally, suppose again that the ith macroblock $\mathbf{x}_i$ is missing. Let $\mathbf{X}_{\partial i}$ denote its neigh-

boring macroblocks and let (m, n) denote the coordinates of the upper left corner of $\mathbf{x}_i$. Establish a search range $\mathcal{S}$ of $(2S + 1) \times (2S + 1)$ pixels in the previous frame X^{-1} centered at (m, n), that is $\mathcal{S} = \{X_{k,l}^{-1} | \ k\epsilon[m - S, m + S], \ l\epsilon[n - S, n + S], \ k, l \text{ integers}\}$. Let $\mathbf{u}$ denote a macroblock sized region in $\mathcal{S}$, that is $\mathbf{u} \subset \mathcal{S}$, and let $\mathbf{u}_B$ denote the boundary pixels of $\mathbf{u}$, then

$$\hat{\mathbf{x}}_{\mathbf{i}} = \arg \max_{\mathbf{u} \subset \mathcal{S}} f(\mathbf{u}_B \mid \mathbf{X}_{\partial i}).$$

Using the potential functions described above this can be written as

$$\hat{\mathbf{x}}_{\mathbf{i}} = \arg \min_{\mathbf{u} \subset \mathcal{S}} \sum_{(r,s)|X_{(r,s)}\epsilon \mathbf{u}_B} \sum_{l=r}^{r+1} \sum_{k=s}^{s+1} \sum_{m=0}^{3} b_{l,k}^m \rho_\gamma \left(\frac{D_m(X_{l,k})}{\sigma} \right). \tag{12}$$

This technique is particularly useful for restoring I frames that have been heavily damaged due to network packet loss.

In the following, we describe how we achieve motion compensated restoration by first estimating the missing motion vectors and then utilizing the estimates of the missing motion vectors to reconstruct the missing macroblocks.

3.4. Motion Compensated: Motion Vector Estimation

Most frames in an MPEG sequence are predicted frames that have motion vectors associated with their macroblocks. Let $\mathbf{v}_i$ be the motion vector associated with the ith macroblock in the current frame. In lossless transmission, $\mathbf{x_i}^{(0)}$, the ith macroblock in the current frame, is reconstructed by the decoder as $\mathbf{x_i}^{(0)} = \mathbf{x_{i-v_i}^{(-1)}} + \mathbf{n}_i$. Here $\mathbf{x_{i-v_i}^{(-1)}}$ is the macroblock in the reference frame that closely matches $\mathbf{x_i}^{(0)}$, $\mathbf{n}_i$ is the error arising from having replaced $\mathbf{x_i}^{(0)}$ by $\mathbf{x_{i-v_i}^{(-1)}}$, and $\mathbf{i}$ indicates the spatial coordinates of the ith macroblock. In lossy transmission it is not possible to recover $\mathbf{n}_i$. The goal is to obtain an estimate for $\mathbf{v}_i$ that will point to $\mathbf{x_{i-v_i}^{(-1)}}$.

One approach to reconstructing a lost macroblock would be to estimate its associated missing motion vector by averaging the motion vectors of surrounding macroblocks [5, 14]. This motion vector estimate is then used to fill in the missing macroblock.

In the following section we present an alternative means of estimating the missing motion vector based on the use of Markov Random Field (MRF) models.

3.4.1. MAP Estimation of Motion Vectors

Utilizing the same MRF model described above to model each component of the motion field we can obtain a MAP estimate of the missing motion vector given its neighboring motion vectors. In addition, using the results of Section 3.2 we immediately infer that the median of the motion vectors of the surrounding macroblocks yields a suboptimal estimate of the missing motion vector. These estimates are then utilized to perform motion compensated restoration of the missing macroblock.

It is also evident from Equation (11), that when $\sigma = 1$, $\gamma \to \infty$ (larger than the image dimensions suffices), $\mathbf{b}_k = 1$ for $k = 1, \cdots, 8$, and z_ks are motion vectors, the MAP estimate of the motion vector is the average of all surrounding motion vectors, as proposed in [5]. Thus, the median and average of the surrounding motion vectors are special cases of the MAP estimate.

3.4.2. Temporal-Spatial Approach

The above described techniques for estimating missing motion vectors rely solely on temporal data (neighboring motion vectors). An alternative approach that uses both temporal and spatial data is described next. The use of both temporal and spatial data yields estimates for missing motion vectors that are more reliable, as will be presented in Section 4.

The alternative approach is based on using a ternary tree to classify the motion vectors neighboring the missing motion vector, and the MRF models of the image and motion field. Each neighboring motion vector is classified into one of nine classes according to whether each of its components are positive, negative, or zero. This is shown in Figure 1 where $\mathbf{v}_x$ denotes the horizontal component of a motion vector and $\mathbf{v}_y$ its vertical component. The idea is to implicitly model the discontinuity in the motion field, and is similar to the approaches considered in [50, 51], wherein the discontinuities of motion fields were modeled by means of binary MRFs.

After classifying all the neighboring motion vectors into their respective classes, we then determine the class or group to which the missing motion vector belongs. This is done by assigning a cost to each class and then choosing the class with the lowest cost. After the class with the lowest cost has been chosen, the motion vectors belonging to that class are

modeled via the MRF and a MAP estimate of the missing motion vector obtained. If the MAP estimate is not unique, the motion vector that provides the macroblock with the "best" matching boundaries, given the neighboring macroblocks, is then chosen.

Formally, let $\mathcal{K} = \{K_i, \ i = 1\cdots 9\}$ denote the set of 9 classes. Also let $\mathbf{v_t}$, $\mathbf{v_b}$, $\mathbf{v_l}$, and $\mathbf{v_r}$, be the top, bottom, left, and right neighboring motion vectors of the missing motion vector $\mathbf{v}_i$, respectively. Then, the cost C_i incurred by assuming that $\mathbf{v}_i \epsilon K_i$ is given by

$$C_i = \alpha[g_i(\mathbf{v}_i, \mathbf{v_t}) + g_i(\mathbf{v}_i, \mathbf{v_b}) + g_i(\mathbf{v}_i, \mathbf{v_l}) + g_i(\mathbf{v}_i, \mathbf{v_r})]$$

where α is some constant and

$$g_i(\mathbf{v}, \mathbf{u}) = \begin{cases} 0 & \mathbf{v}\epsilon K_i, \ \mathbf{u}\epsilon K_i \\ 1 & \text{otherwise,} \end{cases}$$

that is, $g_i(\cdot, \cdot)$ is 0 when both arguments belong to the same class of motion vectors.

Let $\mathcal{K}^*$ denote the set of classes that have the same minimum cost. It is conceivable that two or more classes will have the same cost. In this case we need to use spatial information to decide between which classes of vectors to choose. However, prior to that we need to obtain a MAP estimate of $\mathbf{v}_i$ given that it belongs to the class $K\epsilon\mathcal{K}^*$. This is done by solving the following for every $K\epsilon\mathcal{K}^*$

$$\{\hat{\mathbf{v}}_i\}_K = \arg\max_{\mathbf{v}\epsilon K} f(\mathbf{v} \mid K),$$

where $\{\hat{\mathbf{v}}_i\}_K$ is the class of vectors that maximize the above equation.

Choosing the motion vector with the "best" matching boundaries is performed as follows. Let $\mathcal{V}$ denote the set of MAP motion vectors, that is, $\mathcal{V} = \bigcup_{K\epsilon\mathcal{K}^*}\{\hat{\mathbf{v}}_i\}_K$, then $\hat{\mathbf{v}}_i$, the estimate of $\mathbf{v}_i$, is given by

$$\hat{\mathbf{v}}_i = \arg\max_{\mathbf{v}\epsilon\mathcal{V}} f(\mathbf{B}_{\mathbf{i-v}}^{(-1)} \mid \mathbf{X}_{\partial i}).$$

Here $\mathbf{B}_{\mathbf{i-v}}^{(-1)}$ consists of the pixels lying on the boundary of the macroblock $\mathbf{x}_{\mathbf{i-v}}^{(-1)}$, as illustrated in Figure 2. Using the potential functions described above, this can be rewritten as

$$\hat{\mathbf{v}}_i = \arg\min_{\mathbf{v}\epsilon\mathcal{V}} \sum_{(r,s)|X_{(r,s)}\epsilon\mathbf{B}_{\mathbf{i-v}}^{(-1)}} \sum_{l=r}^{r+1}\sum_{k=s}^{s+1}\sum_{m=0}^{3} b_{l,k}^m \rho_\gamma(\frac{D_m(X_{l,k})}{\sigma}). \tag{13}$$

In Section 4 we will compare the performance of the various motion compensated error

concealment techniques.

4. Experimental Results

To test the reconstruction algorithms, the *salesman*, *football*, *flowergarden*, and hockey sequences, encoded at data rates of 0.3 Mbits/sec, 1.15 Mbits/sec, 1.5 Mbits/sec, and 1.5 Mbits/sec respectively[2], were multiplexed with MPEG1 Layer II audio streams into MPEG1 System Layer Streams. For our experiments we assumed that the protocol for transmitting video over networks is the Asynchronous Transfer Mode (ATM) protocol. In ATM, information is transmitted in fixed size packets of data called "cells". The respective System Layer Streams were then packed into ATM cells. This is done by packing important header information such as MPEG1 system layer or MPEG2 Program Stream pack and packet headers, and any header data necessary for the proper decoding of the compressed sequences, into high priority cells.[3] All other data, such as motion vectors and DCT coefficients, are packed into low priority cells. An extra 16 bits are inserted at the start of each cell, 9 of which provide the location of the first macroblock being packed into the cell, and the rest are used as an ATM cell counter. The extra 9 bits are also used to indicate when a macroblock spans across more than one cell. In addition, for every first macroblock being packed into a cell, the following is performed:

- The address of the macroblock is coded relative to the position of the Slice [9, 10] to which it belongs.

- Any motion vectors associated with the macroblock are coded as is, not differentially.

- If the macroblock is intracoded then the DC coefficients of its first Y block and chrominance blocks are coded relative to a value of 1024.

The above measures are performed for every first macroblock being packed into an ATM cell to prevent the loss of macroblocks in prior cells from affecting the decoding of succeeding

---

[2]The GOPs in each sequence were 15 frames, and the frame pattern of each GOP was IBBPBBPBBPBBPBB

[3]Our goal is to protect the least amount of data and yet be able to reconstruct damaged frames unlike other approaches [13, 14, 15, 16, 20]. Hence we only protect the headers without which proper decoding of the sequence would not be possible.

macroblocks. The packed System Layer Streams were then subjected to 2% and 5% random ATM cell loss.

It is to be noted that coding the address of a macroblock relative to the position of the Slice to which it belongs, absolute coding of its motion vectors, and coding its DC coefficients relative to a value of 1024, are described in the MPEG video standards [9, 10] as the means for encoding the first macroblock of every Slice. This however does not mean that we start a new Slice at the beginning of every ATM cell, as this would result in a larger increase in the overhead than required. Rather, it means that we only utilize the concepts for coding the first macroblock in a Slice. The codes used were those specified by the Huffman coding tables provided in the MPEG standards [9, 10].

To determine which macroblocks were lost due to cell loss, the difference between the addresses of the two most recent correctly decoded macroblocks is obtained. This difference between both addresses is taken to be the number of macroblocks (between both macroblocks) that were lost. This can lead to an error in determining which macroblocks in a P or B frame were actually lost or damaged. This is a consequence of the fact that in MPEG video, predicted macroblocks that have zero motion vectors and negligible difference DCT coefficients are not coded but skipped. Under lossless conditions, a macroblock address difference that is greater than 1 is interpreted by the MPEG decoder to mean that the intervening macroblocks are to be duplicated from the reference frame. In particular, those macroblocks in the reference frame that have the same locations as those that are being decoded are used. Thus, our techniques for estimating which macroblocks are missing can misinterpret skipped macroblocks as being lost. This may result in skipped macroblocks being improperly constructed particularly if all the surrounding macroblocks have non-zero motion vectors.

Having assigned the missing macroblocks, reconstruction proceeds as follows:

- If the damaged frame is an I frame and all the neighbors of a missing macroblock are available, then reconstruction is performed by means of the spatial reconstruction techniques discussed above.

- If the damaged frame is an I frame and some of the neighbors of a missing macroblock are not available, then reconstruction is performed by searching for the macroblock in the most recent I or P frame that optimizes the boundary pixels (Equation (12)). The search space used is 21×21 pixels in size.

- If the damaged frame is a P or B frame, reconstruction as follows:

  - If the missing macroblock is surrounded by intracoded macroblocks it is then reconstructed by the same method used for restoring macroblocks missing from an I frame,

  - otherwise the missing macroblock is assumed to be intercoded and it is reconstructed by first estimating its associated motion vector. This is done by a number of ways enumerated below:

    1. Temporal replacement, that is, a motion vector with zero components is used.
    2. The average of the surrounding motion vectors is obtained.
    3. The median of the surrounding motion vectors is obtained.
    4. The MAP estimate of the missing motion vector given its neighboring motion vectors is obtained.
    5. The Temporal-Spatial approach is used to estimate the missing motion vector.

    The estimate of the missing motion vector is then used to provide error concealment. This is achieved by replacing the missing macroblock by the region in the past I or P reference frame to which the estimated motion vector is pointing. We will compare the performance of all the above enumerated motion vector estimation techniques later on.

For all of the motion compensated techniques, the parameters σ, γ, $\{\mathbf{b}_k, \ k = 1, \cdots, 8\}$ were set to unit value. It was observed that changing the values of these parameters did not significantly impact the quality of the reconstruction.

In Figures 3a and 3b we show the original and damaged frames (due to missing macroblocks), from the *salesman* sequence. Reconstruction was performed spatially in Figure 3c

via median filtering, in Figure 3d by iteratively solving for the MAP estimate of each miss-
ing pixel within the damaged macroblock, and in Figure 3e by means of Bilinear Interpo-
lation [26]. The reconstruction PSNRs were found to be 28.36 dB, 28.14406 dB, and 27.99
dB respectively. From a PSNR point of view the performance of all techniques is compara-
ble. It was observed that the reconstructions due to the first two techniques have sharper
edges. Furthermore, our approach based on median filtering is attractive since it can be
implemented in real-time [52], does not require a search for dominant edges, and operates
on spatial data rather than DCT coefficients which may not be available.

In our experiments it was observed that using values for σ that are greater than one and
$\gamma = 1.0$ provided the best MAP restoration when the weights $\{\mathbf{b}_k , k = 1 \cdots 8\}$ were of unit
value. This can be seen since for $\gamma = 1.0$, $\sigma > 1$ values will not heavily penalize edges. The
image in Figure 3d was reconstructed with $\sigma = 100$. It was also observed that using the
median values of the border pixels as initial values led to rapid convergence to the optimal
MAP estimate.

Figure 4a is a decoded frame from the *salesman* sequence. Due to random cell loss
major portions the frame were lost as shown in Figure 4b. Reconstruction is performed by
minimizing Equation (12). The search space for the motion vectors did not exceed an area
of 7×7 pixels and hence an exhaustive search for the motion vector was implemented. A
small search region was used in this case since there is little motion in this sequence. In the
case of sequences where a substantial amount of motion exists an exhaustive search may be
too costly and thus other searches such as the logarithmic search may be implemented but
at a cost of lower fidelity. As seen the reconstructed version in Figure 4c closely matches the
original in Figure 4a with a reconstruction PSNR value of 35.72 dB. The PSNR value was
obtained via

$$\text{PSNR} = 10 \log \frac{255^2}{MSE(Y) + MSE(U) + MSE(V)}$$

where $MSE(\cdot)$ denotes the mean square error of the reconstructed color component.

Figure 5a is a frame from the *flowergarden* sequence, and Figure 5b is the same frame
with missing macroblocks due to 5% ATM cell loss. The frame is restored via temporal
replacement, obtaining the average of the neighboring motion vectors, finding the median of
the neighboring motion vectors, obtaining the MAP estimate of the missing motion vector,

and via the Temporal-Spatial approach in Figures 5c, 5d, 6a, 6b, and 6c respectively. The reconstruction PSNR values are 26.75 dB, 27.50 dB, 28.77 dB, 30.18 dB, and 30.30 dB respectively.

As is evident, the use of a MRF field that does not penalize the discontinuities in the motion field outperforms using the average or the median of the surrounding motion vectors. It is also observed that using spatial information, as is done in the Temporal-Spatial approach, results in better restoration. This can be seen in Figure 6c where it is evident that the damaged portions of the tree trunk are almost perfectly restored. This is attributed to the fact that the Temporal-Spatial approach attempts to preserve the discontinuities in the motion field while matching the macroblock boundaries. In this case, the "best" matching boundaries where those that preserved the discontinuity between the tree trunk and the background. Also shown in Figures 5c, 5d, 6a, 6b, and 6c are the effect of error propagation due to inaccurate restoration.

For comparison purposes we provide the reconstruction PSNR values for the various error concealment schemes when 5% of the ATM cells were lost, for the *flowergarden*, *football* and *hockey* sequences, respectively in Figures 7, 8 and 9. We also provide the average PSNR values for the different error concealment strategies at 2% and 5% ATM cell loss rate for the same three sequences in Tables 1 and 2, respectively.

Table 1: Average PSNR values in dB for the different error concealment schemes for the *flowergarden*, *football*, and *hockey* sequences at a 2% ATM cell loss rate.

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	32.87	33.98	37.27
Average Motion Vector	35.57	34.54	39.53
Median Motion Vector	35.99	35.09	40.46
MAP estimation of Motion Vector	36.58	35.80	41.21
Temporal-Spatial Approach	36.75	35.80	41.87

In general, restoration based on finding the MAP estimate of the missing motion vector or using the Temporal-Spatial approach was better than that attained by the other techniques by at least 1 dB. Furthermore, the gap in performance between the Temporal-Spatial approach and the MAP estimation of the missing motion vector widened in the case of both the *football* and *hockey* sequences. This is also seen in Figures 8 and 9, and can be attributed

Table 2: Average PSNR values in dB for the different error concealment schemes for the *flowergarden*, *football*, and *hockey* sequences at a 5% ATM cell loss rate.

Error Concealment Technique	*flowergarden*	*football*	*hockey*
Temporal Replacement	30.12	31.51	34.94
Average Motion Vector	32.24	32.41	36.67
Median Motion Vector	32.56	32.83	37.32
MAP estimation of Motion Vector	33.44	33.10	38.10
Temporal-Spatial Approach	33.48	33.60	39.28

to the fact that motion field of *flowergarden* is uniform, unlike that of *football* and *hockey*. In the case of the latter, the motion vectors do not point in one general direction. In addition, it is possible for a macroblock to lie on the boundary between two objects moving in opposite directions. In such a case the use of spatial data was needed to determine which object the missing macroblock belonged to. It is also observed that using the median of the neighboring motion vectors performed better than using the average of the neighboring motion vectors.

5. Conclusion

Due to the nature of how channels errors and congestion are exhibited on data networks, error concealment must be used when video is transmitted. We have presented both spatial and temporal error concealment techniques for compressed video. Our spatial technique based on median filtering, although sub-optimal, can be implemented in real-time. We have also shown that using temporal-spatial data yields better reconstructions than only using temporal data. We are currently investigating the use of DSPs for the deployment of error concealment in video codecs used in ATM networks. Our methods can be extended to wireless networks and streaming video on packet-switched networks, for example, the Internet.

Appendix A.

We show here that $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}|\mathbf{y}=\mathbf{D}\mathbf{x}}[-\ln f(\mathbf{x})]$. Let $\mathcal{A}$ be the event that $\mathbf{y} = \mathbf{D}\mathbf{x}$. In general [53], $f(\mathbf{x}, \mathbf{y})\mathrm{P}(\mathcal{A} \mid \mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y} \mid \mathcal{A})\mathrm{P}(\mathcal{A})$. Using Bayes' rule then,

$$f(\mathbf{x} \mid \mathbf{y})\mathrm{P}(\mathcal{A} \mid \mathbf{x}, \mathbf{y}) = \frac{f(\mathbf{x}, \mathbf{y} \mid \mathcal{A})\mathrm{P}(\mathcal{A})}{f(\mathbf{y})}.$$

But

$$P(\mathcal{A} \mid \mathbf{x}, \mathbf{y}) = \begin{cases} 0 & \mathbf{y} \neq \mathbf{Dx} \\ 1 & \mathbf{y} = \mathbf{Dx} \end{cases}$$

Thus,

$$f(\mathbf{x} \mid \mathbf{y}) = \frac{f(\mathbf{x}, \mathbf{y} \mid \mathcal{A})P(\mathcal{A})}{f(\mathbf{y})}.$$

Let $I_{\delta x} = \{\mathbf{z} \mid \|\mathbf{z} - \mathbf{x}\| \leq \delta x\}$, and $I_{\delta y} = \{\mathbf{u} \mid \|\mathbf{u} - \mathbf{y}\| \leq \delta y\}$. Now,

$$f(\mathbf{x}, \mathbf{y} \mid \mathcal{A}) = \lim_{\delta x \to 0} \lim_{\delta y \to 0} P(\mathbf{x}\epsilon I_{\delta x}, \mathbf{y}\epsilon I_{\delta y} \mid \mathcal{A}).$$

Rewriting $P(\mathbf{x}\epsilon I_{\delta x}, \mathbf{y}\epsilon I_{\delta y} \mid \mathcal{A})$ as $P(\mathbf{x}\epsilon I_{\delta x}, \mathbf{y}\epsilon I_{\delta y} \mid \mathcal{A}) = P(\mathbf{y}\epsilon I_{\delta y} \mid \mathbf{x}\epsilon I_{\delta x}, \mathcal{A})P(\mathbf{x}\epsilon I_{\delta x} \mid \mathcal{A})$, where

$$P(\mathbf{y}\epsilon I_{\delta y} \mid \mathbf{x}\epsilon I_{\delta x}, \mathcal{A}) = \begin{cases} 0 & \mathbf{D}I_{\delta x} \cap I_{\delta y} = \phi \\ 1 & \mathbf{D}I_{\delta x} \cap I_{\delta y} \neq \phi \end{cases}$$

and $\mathbf{D}I_{\delta x} = \{\mathbf{Dz} \mid \mathbf{z}\epsilon I_{\delta x}\}$ then, $f(\mathbf{x}, \mathbf{y} \mid \mathcal{A})P(\mathcal{A}) = f(\mathbf{x} \mid \mathcal{A})P(\mathcal{A})$. Since

$$f(\mathbf{x} \mid \mathcal{A})P(\mathcal{A}) = f(\mathbf{x})P(\mathcal{A} \mid \mathbf{x})$$

then, $L(\mathbf{x} \mid \mathbf{y}) = \ln f(\mathbf{x}) + \ln P(\mathcal{A} \mid \mathbf{x}) - \ln f(\mathbf{y})$. For a given observed image, the third term in the preceding equation is independent of $\mathbf{x}$, hence the MAP estimate is obtained as

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \max_{\mathbf{x}\mid\mathbf{y}=\mathbf{Dx}} [\ln f(\mathbf{x}) + \ln P(\mathcal{A} \mid \mathbf{x})] \\ &= \arg \min_{\mathbf{x}\mid\mathbf{y}=\mathbf{Dx}} [-\ln f(\mathbf{x}) - \ln P(\mathcal{A} \mid \mathbf{x})]. \end{aligned}$$

Since,

$$P(\mathcal{A} \mid \mathbf{x}) = \begin{cases} 0 & \mathbf{y} \neq \mathbf{Dx} \\ 1 & \mathbf{y} = \mathbf{Dx}, \end{cases}$$

then, $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}\mid\mathbf{y}=\mathbf{Dx}} [-\ln f(\mathbf{x})]$.

Appendix B.

We show here that

$$\hat{X}_{i,j} = \frac{\sum_{k=J_1}^{J_2} b_k z_k + \gamma \sigma [\sum_{k=J_2+1}^{8} b_k - \sum_{k=1}^{J_1-1} b_k]}{\sum_{k=J_1}^{J_2} b_k}$$

where J_1 and J_2 satisfy $z_{J_2} - z_{J_1} \le \gamma\sigma$, $J_2 \ge J_1$. We assume positive weights as well as positive values for $z_1 \cdots z_8$.

Consider the following special case

Case2A: $|\boldsymbol{\Delta_k}| > \gamma\sigma \ \forall\mathbf{k}$

In this case the optimal estimate $\hat{X}_{i,j}$ will lie between z_1 and z_8. It will either be within a $\gamma\sigma$ of some z_n or not. If the former is true then, $h_\gamma(\hat{X}_{i,j}) = b_n(\frac{\hat{\Delta}_n}{\sigma})^2 + \sum_{k=1, k\neq n}^{8} b_k[\frac{2\gamma|\hat{\Delta}_k|}{\sigma} - \gamma^2]$ otherwise

$$
\begin{aligned}
h_\gamma(\hat{X}_{i,j}) &= \sum_{k=1}^{8} b_k[\frac{2\gamma|\hat{\Delta}_k|}{\sigma} - \gamma^2] \\
&= \sum_{k=1}^{8} b_k \frac{2\gamma|\hat{\Delta}_k|}{\sigma} - \gamma^2 \sum_{k=1}^{8} b_k
\end{aligned}
$$

where

$$
\hat{\Delta}_k = \begin{cases} z_k - \hat{X}_{i,j} & z_k \epsilon U \\ \hat{X}_{i,j} - z_k & z_k \epsilon L, \end{cases}
$$

Considering the latter, that is $\hat{X}_{i,j} \notin [z_k - \gamma\sigma, z_k + \gamma\sigma] \ \forall k$, let $\tilde{h}_\gamma(X_{i,j}) = \sum_{k=1}^{8} b_k \frac{2\gamma|\hat{\Delta}_k|}{\sigma}$ For the following intervals and for the variable $X_{i,j}$, $\tilde{h}_\gamma(X_{i,j})$ is a straight line with negative, zero, or positive slope.

1. $X_{i,j} < z_1 - \gamma\sigma$:
   $\tilde{h}_\gamma(X_{i,j}) = \frac{2\gamma}{\sigma}[-(\sum_{k=1}^{8} b_k)X_{i,j} + \sum_{k=1}^{8} b_k z_k]$

2. $z_l - \gamma\sigma < X_{i,j} < z_{l+1} - \gamma\sigma$ for $l = 1\cdots 7$:
   $\tilde{h}_\gamma(X_{i,j}) = \frac{2\gamma}{\sigma}[(\sum_{k=1}^{l} b_k - \sum_{k=l+1}^{8} b_k)X_{i,j} + \sum_{k=l+1}^{8} b_k z_k - \sum_{k=1}^{l} b_k z_k]$

3. $z_8 - \gamma\sigma < X_{i,j}$:
   $\tilde{h}_\gamma(X_{i,j}) = \frac{2\gamma}{\sigma}[(\sum_{k=1}^{8} b_k)X_{i,j} - \sum_{k=1}^{8} b_k z_k]$.

Since $\tilde{h}_\gamma(X_{i,j})$ is continuous, then there exists an integer J such that $\sum_{k=1}^{J-1} b_k - \sum_{k=J}^{8} b_k < 0$, $\sum_{k=1}^{J} b_k - \sum_{k=J+1}^{8} b_k = 0$, and $\sum_{k=1}^{J+1} b_k - \sum_{k=J+2}^{8} b_k > 0$ that is $\tilde{h}_\gamma(X_{i,j})$ has a flat bottom, or $\sum_{k=1}^{J-1} b_k - \sum_{k=J}^{8} b_k < 0$ and $\sum_{k=1}^{J} b_k - \sum_{k=J+1}^{8} b_k > 0$. If $\tilde{h}_\gamma(X_{i,j})$ does have a flat bottom, then $z_J \le \hat{X}_{i,j} \le z_{J+1}$, otherwise $\hat{X}_{i,j} = z_J$.

Case2B: $|\boldsymbol{\Delta_k}| \le \gamma\sigma$ for some $\mathbf{k}$

Suppose now that there exist J_1 and J_2 such that $J_1 < J_2$ and $z_{J_2} - z_{J_1} \leq \gamma\sigma$, then for $z_{J_1} \leq X_{i,j} \leq z_{J_2}$,

$$h_\gamma(X_{i,j}) = \sum_{k=1}^{J_1-1} b_k[\frac{2\gamma}{\sigma}|X_{i,j} - z_k| - \gamma^2] + \sum_{k=J_2+1}^{8} b_k[\frac{2\gamma}{\sigma}|X_{i,j} - z_k| - \gamma^2] + \sum_{k=J_1}^{J_2} b_k(\frac{X_{i,j} - z_k}{\sigma})^2.$$

In light of **Case 2A** above, if $\sum_{k=1}^{J_1-1} b_k - \sum_{k=J_1}^{8} b_k > 0$, then there exists an integer J_3 such that $J_3 \leq J_1$ and $z_{J_3} \leq \hat{X}_{i,j} \leq z_{J_3+1}$, or $\hat{X}_{i,j} = z_{J_3}$. Similarly, if $\sum_{k=1}^{J_2} b_k - \sum_{k=J_2+1}^{8} b_k < 0$, then there exists an integer J_4 such that $J_4 \geq J_2$ and $z_{J_4} \leq \hat{X}_{i,j} \leq z_{J_4+1}$ or $\hat{X}_{i,j} = z_{J_4}$. Otherwise, there exists an integer J_5 such that $J_1 \leq J_5 \leq J_2$ and $z_{J_5} - \gamma\sigma \leq \hat{X}_{i,j} \leq z_{J_5} + \gamma\sigma$. Since,

$$h_\gamma(X_{i,j}) = \sum_{k=1}^{J_1-1} b_k[\frac{2\gamma}{\sigma}|X_{i,j} - z_k| - \gamma^2] + \sum_{k=J_2+1}^{8} b_k[\frac{2\gamma}{\sigma}|X_{i,j} - z_k| - \gamma^2] + \sum_{k=J_1}^{J_2} b_k(\frac{X_{i,j} - z_k}{\sigma})^2$$

$$\sum_{k=1}^{J_1-1} b_k[\frac{2\gamma}{\sigma}[X_{i,j} - z_k] - \gamma^2] + \sum_{k=J_2+1}^{8} b_k[\frac{2\gamma}{\sigma}[z_k - X_{i,j}] - \gamma^2] + \sum_{k=J_1}^{J_2} b_k(\frac{X_{i,j} - z_k}{\sigma})^2$$

$$\frac{\partial}{\partial X_{i,j}} h_\gamma(\mathbf{x}) = 0, \Rightarrow$$

$$\hat{X}_{i,j} = \frac{\sum_{k=J_1}^{J_2} b_k z_k + \gamma\sigma[\sum_{k=J_2+1}^{8} b_k - \sum_{k=1}^{J_1-1} b_k]}{\sum_{k=J_1}^{J_2} b_k}$$

where J_1 and J_2 satisfy $z_{J_2} - z_{J_1} \leq \gamma\sigma$, $J_2 \geq J_1$. As $\gamma\sigma \to \infty$, **Case 2A** above is satisfied, that is, there exits an integer J such that $z_J \leq \hat{X}_{i,j} \leq z_{J+1}$ when $\sum_{k=1}^{J} b_k = \sum_{k=J+1}^{8} b_k$, or $\hat{X}_{i,j} = z_J$ when $\sum_{k=1}^{J} b_k \geq \sum_{k=J+1}^{8} b_k$. If $J = 3$, then a possible value for $\hat{X}_{i,j}$, would be the median of $z_1 \cdots z_8$. This is in fact the case when all the weights $b_1 \cdots b_8$ are equal.

3. REFERENCES

[1] M. Tomordy, "Airline with the personal touch," *IEE Review*, vol. 44, no. 6, pp. 261–264, November 1998.

[2] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society,* series B, vol. 36, pp. 192–326, 1974.

[3] R. Kinderman and J. L. Snell, *Markov Random Fields and their Applications*, vol. 1 of *Contemporary Mathematics.* American Mathematical Society, 1980.

[4] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, November 1984.

[5] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE Journal on Selected Areas in Communication*, vol. 7, no. 5, pp. 807–814, June 1989.

[6] H. Man, F. Kossentini, and M. J. T. Smith, "A class of EZW image coders for noisy channels," *Proceedings of the International Conference on Image Processing*, vol. III, October 26–29 1997, Santa Barbara, California, pp. 90–93.

[7] J. Hagenauer, "Rate compatible punctured convolutional codes (RCPC Codes) and their applications," *IEEE Transactions on Communications*, vol. 36, no. 4, pp. 389–400, April 1988.

[8] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–996, May 1998.

[9] ISO/IEC 11172–2 MPEG–1 Video Coding Standard, *ISO/IEC 11172–2, Information Technology–coding of moving pictures and associated audio for digital storage media at up to about 1. 5 Mbits/s–Part2: Video*. ISO, 1993.

[10] ISO/IEC 13818–2 MPEG–2 Video Coding Standard, *ISO/IEC 13818–2, Generic Coding of Moving Pictures and Associated Audio Information–Part2: Video*. ISO, 1995.

[11] ITU-T, *CCIR Recommendation H. 261: Codec for audiovisual services at px64kbits/sec*, 1990.

[12] ITU-T, *Draft ITU-T Recommendation H.263 Version 2: Video Coding for Low Bitrate Communication*, September 1997.

[13] F. Kishino, K. Manabe, Y. Hayashi, and H. Yasuda, "Variable bit rate coding of video signals for ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, pp. 801–806, June 1989.

[14] M. Ghanbari and V. Seferidis, "Cell-loss concealment in ATM video codecs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 3, pp. 238–247, June 1993.

[15] M. Ghanbari and C. J. Hughes, "Packing coded video signals into ATM cells," *IEEE/ACM Transactions on Networking*, vol. 1, no. 5, pp. 505–508, October 1993.

[16] P. Pancha and M. El Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Communications Magazine*, vol. 32, no. 5, pp. 54–66, May 1994.

[17] W. Luo and M. El Zarki, "Analysis of error concealment schemes for MPEG-2 video transmission over ATM based networks," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 1605, May 1995, Taipei, Taiwan, pp. 1358–1368.

[18] L. H. Kieu and K. N. Ngan, "Cell loss concealment techniques for layered video codec in an ATM network," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 666–677, September 1994.

[19] D. Raychaudhuri, H. Sun, and R. S. Girons, "ATM transport and cell-loss concealment techniques for MPEG video," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, November 1993, Minneapolis, Minnesota, pp. 117–120.

[20] C. Hahm and J. Kim, "An adaptive error concealment in SNR scalable system," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 2501/3, May 24–26 1995, Taipei, Taiwan, pp. 1380–1387.

[21] U. Black, *ATM:Foundation for broadband networks*. Prentice Hall Series in Advanced Communications Technology, Prentice Hall, 1995.

[22] V. Parthasarathy, J. Modestino, and K. S. Vastola, "Design of a transport coding scheme for high quality video over ATM networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 358–376, April 1997.

[23] A. S. Tom, C. L. Yeh, and F. Chu, "Packet video for cell loss protection using deinterleaving and scrambling," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 1991, Toronto, Canada, pp. 2857–2860.

[24] Q. Zhu, Y. Wang, and L. Shaw, "Coding and cell loss recovery in DCT based packet video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 3, pp. 248–258, June 1993.

[25] J. Y. Park, M. H. Lee, and K. J. Lee, "A simple concealment for ATM bursty cell loss," *IEEE Transactions on Consumer Electronics*, vol. 39, no. 3, pp. 704–710, August 1993.

[26] P. Salama, N. Shroff, E. J. Coyle, and E. J. Delp, "Error concealment techniques for encoded video streams," *Proceedings of the International Conference on Image Processing*, vol. I, October 23–26 1995, Washington, DC, pp. 9–12.

[27] P. Salama, N. Shroff, and E. J. Delp, "A bayesian approach to error concealment in encoded video streams," *Proceedings of the International Conference on Image Processing*, vol. II, September 16–19 1996, Lausanne, Switzerland, pp. 49–52.

[28] P. Salama, N. Shroff, and E. J. Delp, "A fast suboptimal approach to error concealment in encoded video streams," *Proceedings of the International Conference on Image Processing*, vol. II, October 26–29 1997, Santa Barbara, California, pp. 101–104.

[29] P. Salama, N. B. Shroff, and E. J. Delp, "Error concealment in encoded video streams," in *Image Recovery Techniques for Image Compression Applications*, Ed. N. P. Galatsanos and A. K. Katsaggelos, Kluwer Publishers, 1998.

[30] Y. Wang, Q. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Transactions on Communications*, vol. 41, no. 10, pp. 1544–1551, October 1993.

[31] Y. Wang and Q. Zhu, "Signal loss recovery in DCT-based image and video codecs," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 2501/3, November 1991, Boston, Massachusetts, pp. 667–678.

[32] L. T. Chia, D. J. Parish, and J. W. R. Griffiths, "On the treatment of video cell loss in in the transmission of motion-JPEG and JPEG images," *Computers and Graphics: Image Communication*, vol. 18, no. 1, pp. 11–19, January-February 1994.

[33] H. Sun and J. Zdepski, "Adaptive error concealment algorithm for MPEG compressed video," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 1818, November 1992, Boston, Massachusetts, pp. 814–824.

[34] W. Kwok and H. Sun, "Multidirectional interpolation for spatial error concealment," *IEEE Transactions on Consumer Electronics*, vol. 3, no. 39, pp. 455–460, August 1993.

[35] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 470–477, April 1995.

[36] S. Aign and K. Fazel, "Error detection and concealment measures in MPEG–2 video decoder," *Proceedings of the International Workshop on HDTV*, October 1994, Torino, Italy.

[37] S. Aign and K. Fazel, "Temporal and spatial error concealment techniques for hierarchical MPEG–2 video codec," *Proceedings of the International Workshop on Communications*, June 18–20 1995, Seattle, Washington, pp. 1778–1783.

[38] S. Aign, "Error concealment enhancement by using the reliability outputs of a SOVA in MPEG–2 video decoder," *Proceedings of the URSI International Symposium on Signal, Systems, and Electronics*, October 25–27 1995, San Francisco, California, pp. 59–62.

[39] J. F. Shen and H. M. Hang, "Compressed image concealment and postprocessing for digital video recording," *Proceedings of the IEEE Asia-Pacific Conference on Circuits and Systems*, December 5–8 1994, Taipei, Taiwan, pp. 636–641.

[40] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 - theory," *IEEE Transactions on Medical Imaging*, vol. MI-1, no. 2, pp. 81–94, October 1982.

[41] R. Talluri, "Error resilient video coding in the ISO MPEG-4 standard," *IEEE Communications Magazine*, vol. 36, no. 6, pp. 112–119, June 1998.

[42] J. Liang and R. Talluri, "Tools for robust image and video coding in JPEG2000 and MPEG4 standards," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 3653, January 23–29 1999, San Jose, California, pp. 40–51.

[43] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Transactions on Systems Man and Cybernetics*, vol. 24, no. 3, pp. 455–469, March 1994.

[44] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–382, March 1992.

[45] C. Bouman and K. Sauer, "A generalized gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[46] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76–89, March 1987.

[47] R. Schultz and R. L. Stevenson, "A bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 233–241, May 1994.

[48] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 1981.

[49] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, series B, vol. 48, no. 3, pp. 259–302, 1986.

[50] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 910–926, September 1992.

[51] J. Li, X. Lin, and C. C. J. Kuo, "Boundary control vector motion field representation and estimation by using a markov random field model," *Journal of Visual Communication and Image Representation*, vol. 7, no. 3, pp. 230–243, September 1996.

[52] E. Asbun and E. J. Delp, "Real-time error concealment in compressed digital video streams," *Proceedings of the Picture Coding Symposium*, April 21–23 1999, Portland, Oregon.

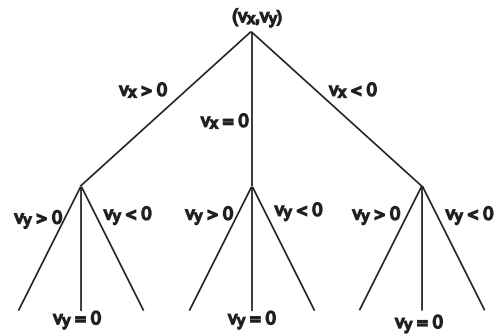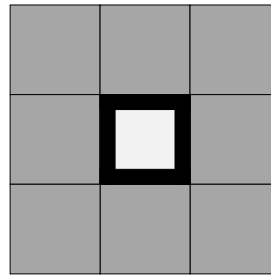[53] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* McGraw Hill, third ed., 1984.

Figure 1: Tree classification of motion vectors.

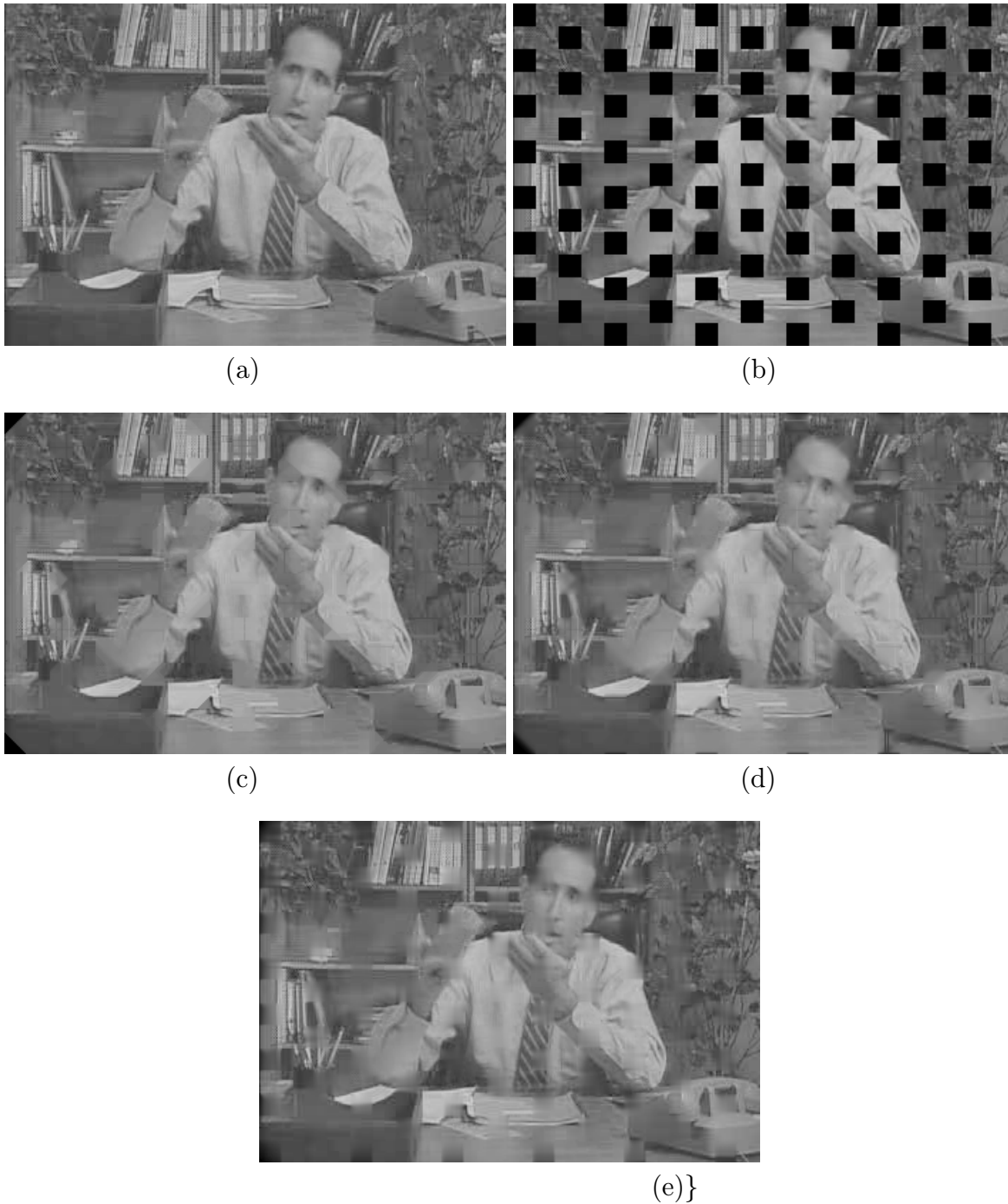Figure 2: Boundary pixels of prospective macroblock

.

Figure 3: Spatial reconstruction. (a) decoded frame from the *salesman* sequence, (b) same frame with missing macroblocks, (c) reconstructed using median filtering, (d) reconstructed by using line search techniques to obtain the MAP estimates with $\sigma = 100.0$, $\gamma = 1.0$, $\mathbf{b}_k = 1.0$ for $k = 1 \cdots 8$, (e) reconstructed using bilinear interpolation.

(a)   (b)

(c)
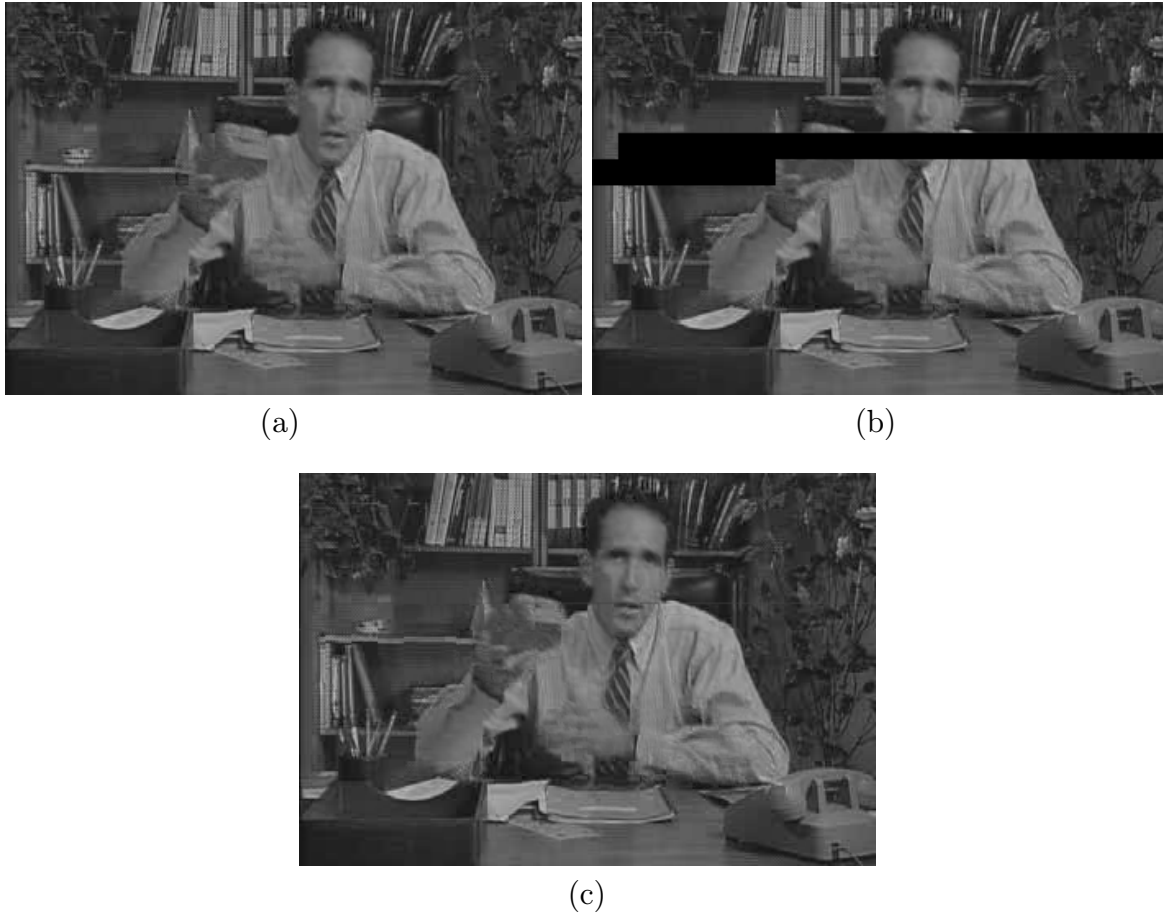
Figure 4: Reconstruction based on finding the macroblock that has "best" matching boundaries. (a) decoded frame from the *salesman* sequence, (b) damaged version due to cell loss, and (c) the reconstructed version.
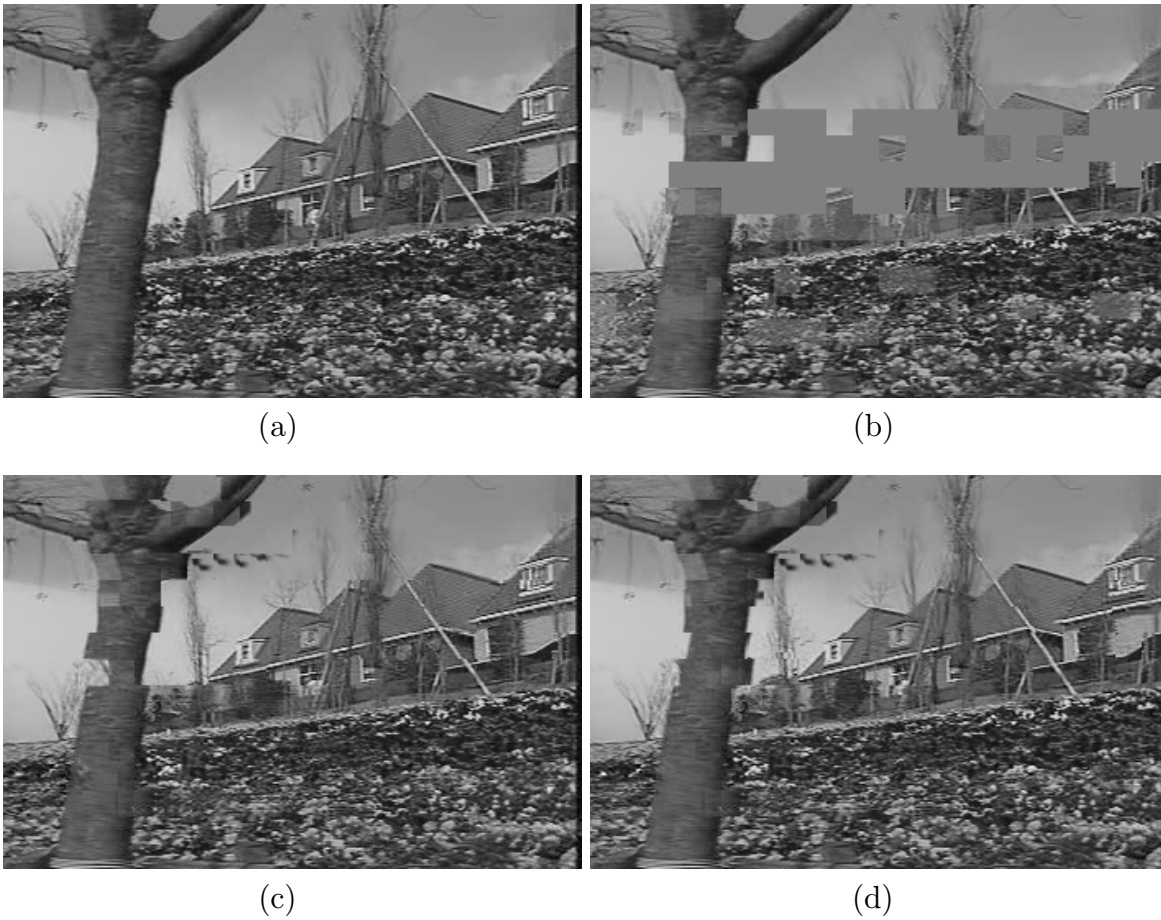
Figure 5: (a): decoded frame from the *flowergarden* sequence, (b): frame is damaged due to 5% ATM cell loss, (c): the frame was restored by using temporal replacement, (d): the frame was reconstructed by finding the average of the neighboring motion vectors. The PSNR values are 26.75 dB, 27.50 dB, and 28.77 dB respectively.

(a)

(b)

(c)

Figure 6: (continuation of previous figure) (a): the frame was restored by finding the median of the neighboring motion vectors, (b): the frame was reconstructed by finding the MAP estimate of the missing motion vector, and (c): the frame was restored by using the Temporal-Spatial approach. The PSNR values are 28.77 dB, 30.18 dB, and 30.30 dB respectively.
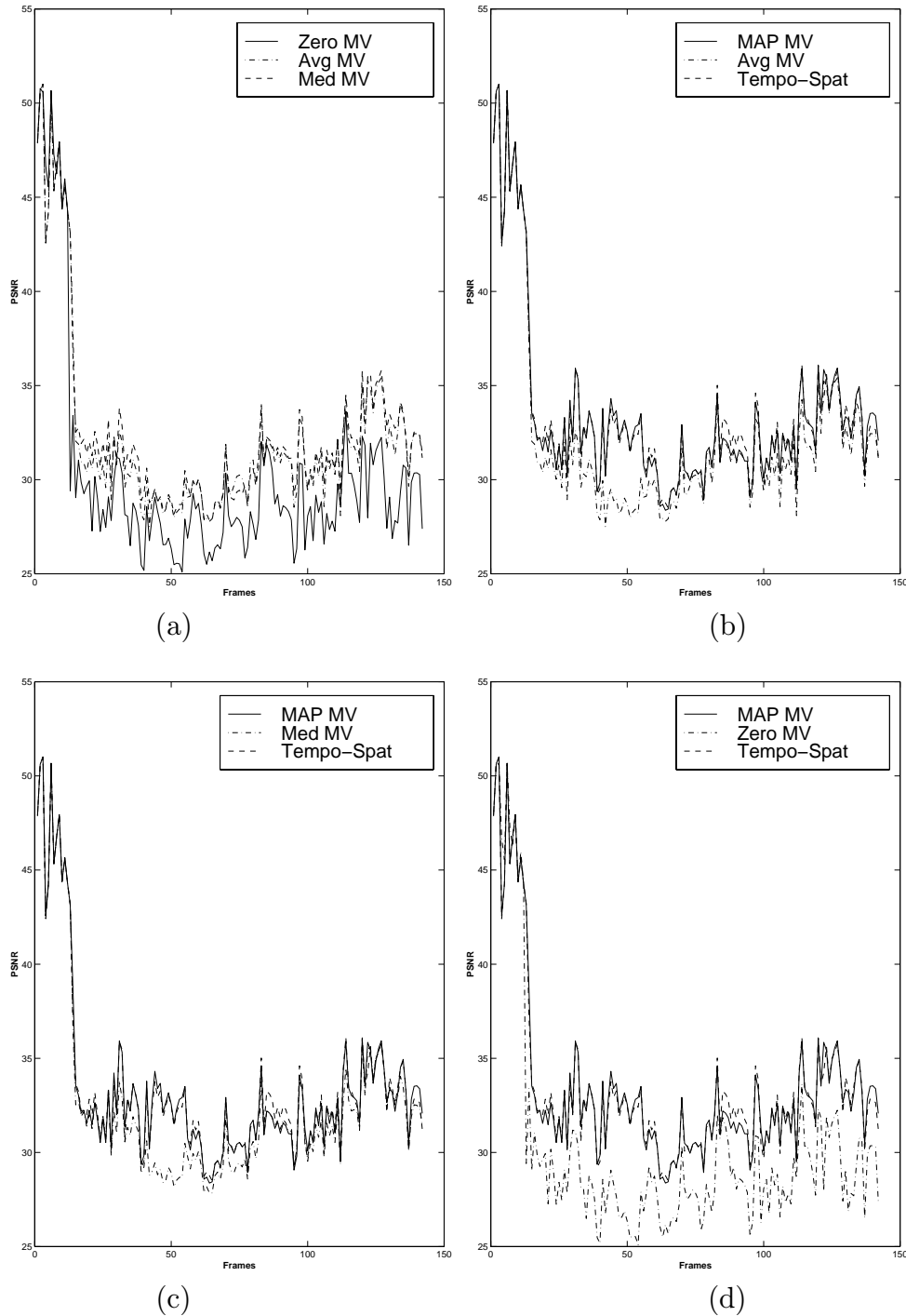
Figure 7: Reconstruction PSNR values for the *flowergarden* sequence when 5% of the cells were dropped. (a): zero motion vector (temporal replacement), the average of the neighboring motion vectors, and the median of the neighboring motion vectors used. (b): the MAP estimate of the missing motion vector, the temporal spatial approach, and the average of the neighboring motion vectors are used. (c): the MAP estimate of the missing motion vector, the temporal spatial approach, and the median of the neighboring motion vectors are used. (d): the MAP estimate of the missing motion vector, the Temporal-Spatial approach, and temporal replacement are used.
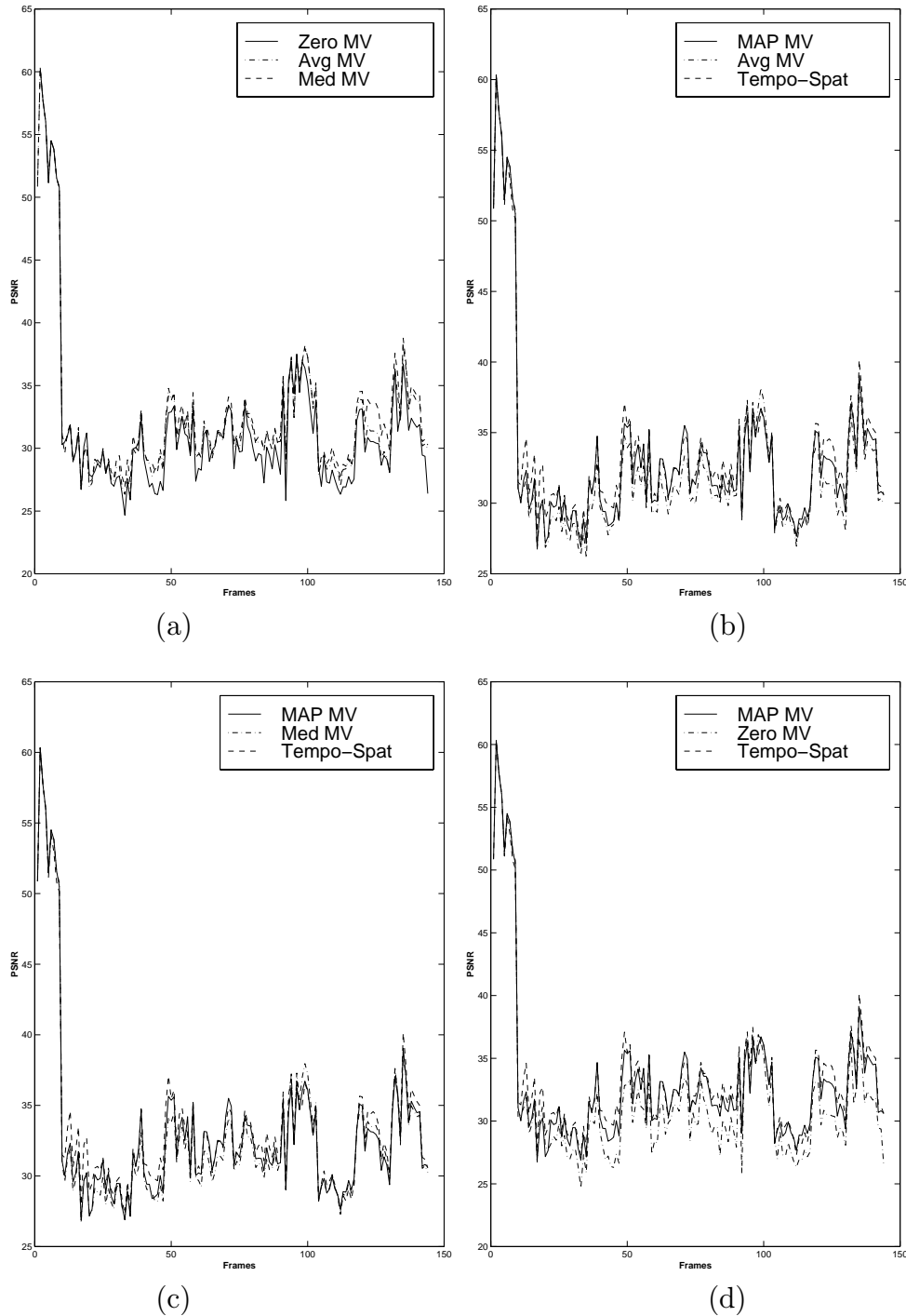
Figure 8: Reconstruction PSNR values for the *football* sequence when 5% of the cells were dropped. (a): zero motion vector (temporal replacement), the average of the neighboring motion vectors, and the median of the neighboring motion vectors used. (b): the MAP estimate of the missing motion vector, the temporal spatial approach, and the average of the neighboring motion vectors are used. (c): the MAP estimate of the missing motion vector, the temporal spatial approach, and the median of the neighboring motion vectors are used. (d): the MAP estimate of the missing motion vector, the temporal spatial approach, and temporal replacement are used.
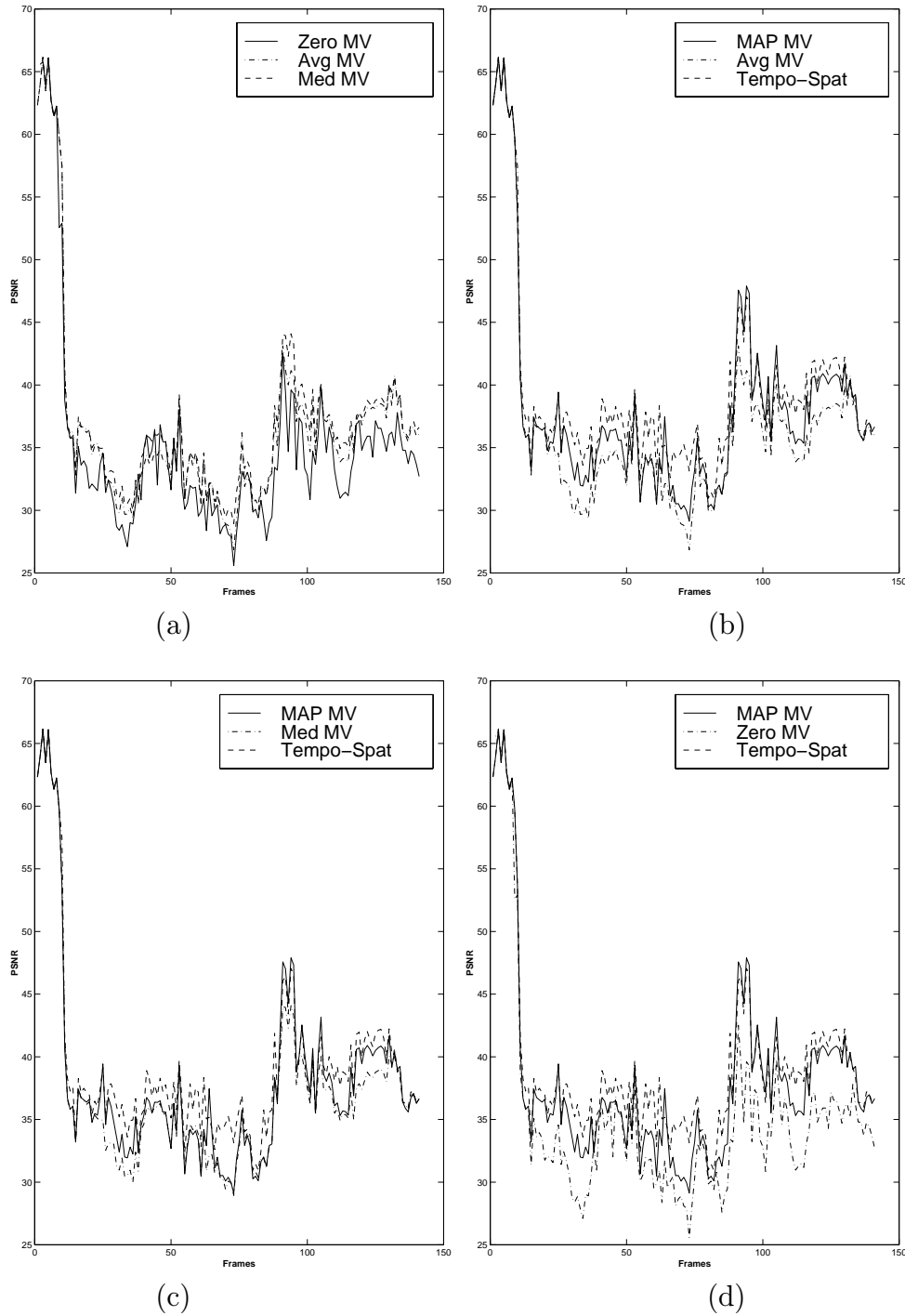
Figure 9: Reconstruction PSNR values for the *hockey* sequence when 5% of the cells were dropped. (a): zero motion vector (temporal replacement), the average of the neighboring motion vectors, and the median of the neighboring motion vectors used. (b): the MAP estimate of the missing motion vector, the temporal spatial approach, and the average of the neighboring motion vectors are used. (c): the MAP estimate of the missing motion vector, the temporal spatial approach, and the median of the neighboring motion vectors are used. (d): the MAP estimate of the missing motion vector, the temporal spatial approach, and temporal replacement are used.