私立東海大學

資訊工程與科學研究所

碩士論文

MPEG 視訊壓縮之錯誤隱蔽技術

Temporal Error Concealment for MPEG Video Compression

研 究 生：練秀佾

指導教授：黃育仁

中華民國九十四年一月

摘要

經過壓縮過後的影像序列在錯誤率高的網路傳輸時, 所遺失的資料經常導致影像被嚴重的破壞，因此需要有效率的錯誤隱蔽處理來減少傳送錯誤的衝擊。目前已經有許多錯誤隱蔽技術提出，不過，當影像序列中的物體的移動過於快速或複雜時，這些錯誤隱蔽技術總是效率不彰。在本論文中，我們發展出二個適合在動態影像壓縮標準 MPEG 的時間錯誤隱蔽演算法來解決此問題。首先，本研究將每一個損壞區塊被切割成四個相同尺寸的子區塊，再使用邊界匹配演算法，利用損壞區塊的周圍未損壞區塊的運動向量和像素的訊息來預測失去的運動向量以重建損壞的區塊，但發現需要大量的計算，所以再提出另一個解決方法；另一個方法是在空間域使用類神經網路模型中之自我組織映射圖網路(self-organizing map, SOM)進行隱蔽錯誤，SOM 使用盲目聚類方式來達到分群目的，因此 SOM 在本研究中用於估計和重組損壞區塊的運動向量。由實驗結果得知，本研究所發展的錯誤隱蔽演算法在視覺品質和客觀評估上比起傳統的方式有重大的改進, 因此本研究對於容易出現錯誤的網路將十分有幫助。

關鍵字： 錯誤隱蔽、類神經網路、自我組織映射圖、運動向量估計、動態影像壓縮標準

Abstract

When transmitted over error prone networks, the compressed video can suffer severe degradation. An efficient error concealment (EC) scheme is essential for diminishing the impact of transmission errors in a compressed video. A number of EC techniques have been developed to combat the transmission errors. However, the techniques are always inefficient when the motions of object in a video are fast or complex. In this thesis, we propose two adaptive temporal EC algorithms to conceal the errors for MPEG-coded video. In the first method, each damaged macroblock is further divided into four sub-blocks with equal size. The information of undamaged motion vector and pixels surrounding a damaged macroblock is used to estimate the lost motion vectors of the sub-blocks based on the boundary matching technique. The estimated motion vectors are used to reconstruct the damaged macroblock by exploiting the information in reference frame. However, this approach entails a considerable amount of processing complexity at the decoder. Thus, we propose another method perform high computational efficiency and good visual quality. The second method is an adaptive EC algorithm that conceals the error for macroblock-based coding systems by using neural network (NN) techniques in the spatial domain. In the proposed algorithm, the self-organizing map (SOM) is used to estimate and reconstruct the lost motion vectors of damaged blocks. The SOM has a great capacity for visualizing and interpreting high-dimensional data sets. Simulation results show that the visual quality and the PSNR evaluation of reconstructed frames are significantly improved by using the proposed EC algorithms. From the experimental results, we find that the proposed algorithms are expected to be useful EC algorithms for motion vector compressed video in error-prone networks.

Keywords： error concealment, neural network, SOM, motion vector estimation, MPEG

Index

List of Tables

List of Figures

Chapter 1

Introduction

In the last decade, a great effort has been spent in studying effective solutions for transmitting digital videos over data networks. In particular, packet video is the emerging technology for a wide range of applications and services, including video conferencing and digital TV over the Internet. The quality of the video data delivered to the end user was a crucial point for the overall performance of these applications. Standard compression techniques, e.g., MPEG-1, MPEG-2 and MPEG-4 [1], were usually adopted to reduce the needs in terms of bandwidth, while ensuring an acceptable reconstructing quality. Transmission of compressed video over an error prone network, such as Internet, may cause information losses due to congestion. The losses further results serious error propagation especially for compressed video that based on motion compensation. Many techniques have been developed to combat the effects of channel error on transmitted video [2]. However, the residual transmission errors were unavoidable. Thus error concealment (EC) techniques always formed the last line of defense for eliminating the visual degradation.

There were two main catalogs of EC methods, i.e. spatial EC and the temporal EC techniques. The spatial EC methods recovered the errors by exploiting the spatial information in a frame. The temporal EC methods utilized the temporal information in a video sequence to reconstruct damaged macroblocks. Generally, temporal EC methods are much practical for video and its complexity is less than that of spatial EC methods [3-4]. Hence this study concentrates on the improvement of temporal EC.

The temporal EC techniques use a macroblock in the reference frame to reconstruct the damaged macroblock. This strategy performs very well for concealing errors in a video. However, motion vectors may be lost at a high packet loss. The motion vector reconstruction stage played a significant role in the temporal concealment procedure. In the past years, various approaches have been proposed to reconstruct lost motion vectors. We notice that the conventional temporal EC methods utilize different linear statistical relationships to conceal damaged macroblocks. If the motions of macroblocks in frames are irregular, the implementation will become inefficient. Thus, we desire to develop simple and flexible temporal EC schemes to solve the problem.

The artificial neural network (NN) techniques have been applied to solve complex problems in the fields of image processing. The self-organizing map (SOM) [5] is an effective neural network model for the visualization of high-dimensional data. The SOM learning algorithm is a type of unsupervised, iterative vector quantization that converts complex, nonlinear statistical data items from a high-dimensional space onto simple reference vectors in a low-dimensional space [6-10]. This study employs the SOM as a predictor to estimate the motion vector of the damaged macroblock. The estimation model proposed herein can exploit the nonlinearity property of the SOM to estimate lost motion vectors more accurately.

To devise a better technique, various hybrid algorithms have been proposed [11-17]. The conventional hybrid algorithms have been shown a poor performance for video sequences with scene changes or complex motion due to false coding mode estimation. A precise motion vector estimation method is essential for a robust temporal EC algorithm. Thus this study proposed sophisticated temporal EC methods to solve the problem. The proposed method aims to produce more precise recovered motion vector for a damaged macroblock.

In Chapter 2, the concept of video compression, the error resilience and quality evaluation are introduced. The EC algorithm using the sub-block technique and experimental results was represented in Chapter 3. Chapter 4 describes the proposed SOM model for EC technique, and demonstrates the result of experiments. Finally, the conclusion and discussion were given in Chapter 5.

Chapter 2

Reviews of Video compression and Error Concealment

The MPEG-coded video stream is very sensitive to the channel disturbances. With some types of network there is a relatively high probability that transmission errors will be present in the bit-stream received by the decoder. A single bit error in the bit-stream may cause a high degradation of image quality due to error propagation. Hence, for frame reconstruction using EC techniques may be required at the receiver. For broadcast application in order to assume graceful degradation as in analog system, MPEG hierarchical coding profile will be needed. To avoid degradation or to enhance the image quality, one may apply the EC techniques at the receiver [1, 11].

2.1 Video Compression

In the context of compression, since video is simply a sequence of digitized images, video is also refer to as moving pictures and the terms “frames” and “picture” are used interchangeably. In principle, one approach to compressing a video source is to apply the JPEG coding to each frame independently. This approach is known as motion JEPG (MJPEG).

In practice, in addition to the spatial redundancy presents in each frame, considerable redundancy is often present between a set of frames since only a small portion of each frame is involved with any motion that is taking place. In the case of a person moving across the screen in a movie, for examples, since a typical scene in a movie has a minimum duration if about 3 seconds, assuming a frame refresh rate if 60 frames per second, each scene is composed of a minimum of 180 frames. Hence by sending only information relating to those segments of each frame that have movement associated with them, considerable additional saving bandwidth can be made by exploiting the temporal differences that exist between many of the frames.

The detailed description is listed below:

(1) Original video sequence (as shown in Fig. 2.1)**.**

- Video frames are independent to each other.
- The considerable spatial redundancy present in each frame.

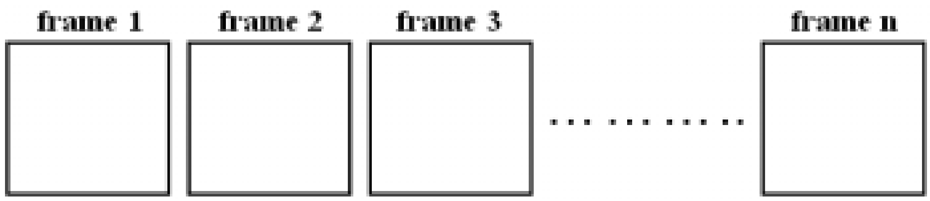


Fig. 2.1 Original video sequence

(2) Video Compression standard：MPEG1, MPEG2, MPEG4, H.263, etc.

- Some frames are referred to others (as shown in Fig. 2.2).
- Only information relating to movement associated with each frame are sent.
- Exploiting the temporal differences that exist between many of the frames can make considerable additional saving bandwidth.

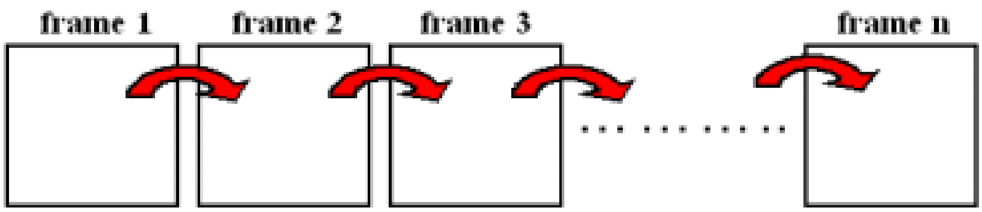


Fig. 2.2 Compressed video sequence

(3) Video coding mode (as shown in Fig. 2.3).

There are two basic types of compressed frame: those that are encoded independently and those are predicted. The first are known as intra-coded frames (I-frames). There are two types of predicted frames: predictive frames (P-frames) and bidirectional frames (B-frames). Because of the way they are derived, the latter are also known as interceded or interpolation frames.

- I-frame: encoded independently
- P-frame: relative to the contents of either a preceding I-frame or a preceding P-frame
- B-frame: predicted using search regions in both past and further frames.

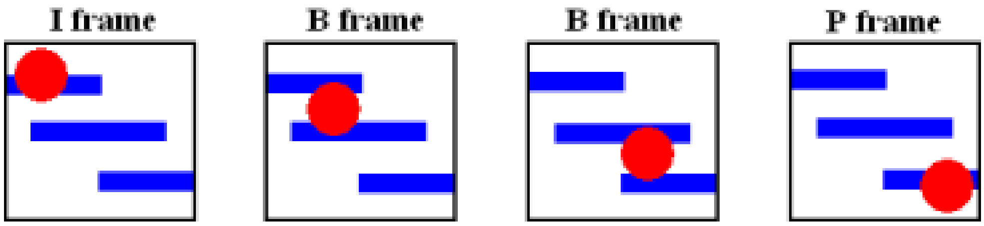


Fig. 2.3 Video coding modes

(4) Error propagation problems (as shown in Fig. 2.4).

Due to video compression features the search regions of predicted frame is based on previous or further reference frames. For example, if the B-frame is predicted by P-frame, then the error in P-frame might be propagated to whole B-frame. Because of these error propagation problems, the error resilient encoding and EC techniques are necessity.

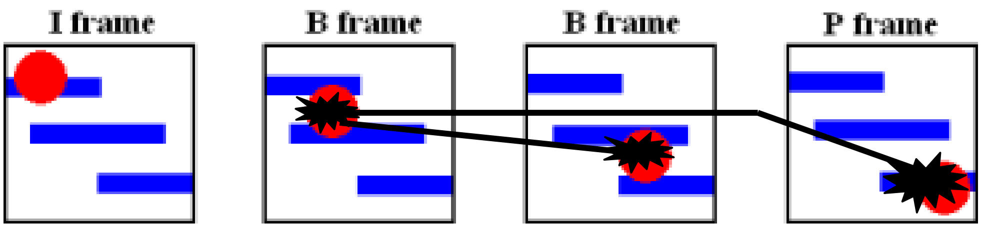


Fig. 2.4 Error propagation problems

2.2 Error Resilience

Here we briefly describe general techniques that have been developed for error resilient (ER) video coding. ER techniques can be divided into three categories, depending on the role that the encoder, decoder, or the network layer plays in the process. Mechanisms devised for combating transmission errors can be categorized into three groups:

(1) Error resilient encoding (forward technique)

(2) Decoder EC (post processing)

(3) Encoder and decoder interactive error control

In this thesis the EC technique is the main theme. Later the detailed description and simulation will be present.

2.2.1 EC Methodology

EC at decoder refers to the recovery or estimation of lost information due to transmission error. Given the block-based hybrid coding paradigm, there are three types of information that may need to be estimated in a damaged macroblock. The first one is the texture information, including the pixel or DCT coefficient values for either an original image macroblock or prediction error macroblock. The second is the motion information, concerning of motion vectors (MVs) for macroblock coded in either P- or B-frame. Third is the coding mode of the macroblock. This mechanism is a post processing that adds to the decoder. There are two kinds of methods, i.e. spatial and temporal domain EC techniques. Here, the temporal domain EC techniques are topic of this study.

2.2.2 Spatial Domain EC

The spatial EC exploits the spatial redundancy in one frame. This technique is proposed for intra-coded frames, where no motion information exists. Spatial EC always assumes that variation of pixels is small in an image. Based on spatial domain EC technique, the linear interpolation algorithm can reconstruct lost image content, shown in Fig. 2.5. However, there is blurring in the edge region by spatial domain EC technique.

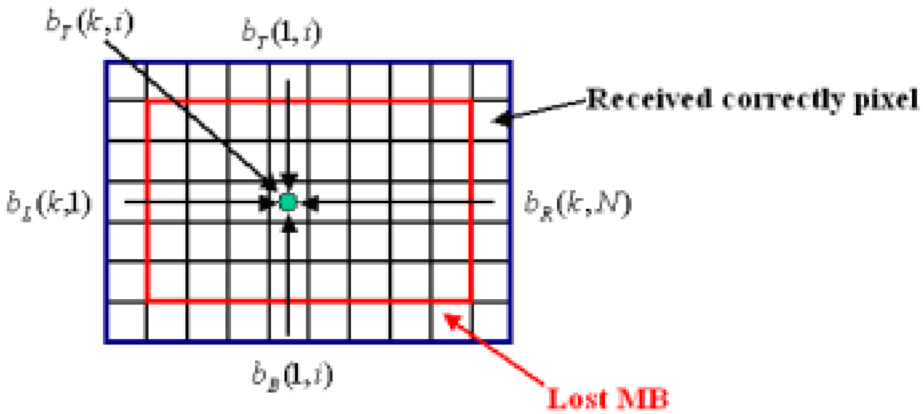


Fig. 2.5 Linear interpolation algorithm

In the spatial domain EC with linear interpolation algorithm, the lost macroblock is reconstructed with the received correctly pixels. Every constructed pixel in the lost macroblock is decomposition of four components (left, right, top and bottom). Four directional components are inverse proportion of distance. The linear interpolation can be defined as

$$b(i,k)=\frac{\left[d_R b_L(k,1)+d_L b_R(k,n)+d_B b_T(1,i)+d_T b_B(n,i)\right]}{d_L+b_R+d_T+b_B}, \qquad (1)$$

where d_L is the distance between $b_L(k,1)$ and $b(i,k)$; d_R is the distance between $b_R(k,n)$ and $b(i,k)$; d_T is the distance between $b_T(1,i)$ and $b(i,k)$; d_B is the distance between $b_B(n,i)$ and $b(i,k)$.

2.2.3 Temporal Domain EC

Temporal domain EC makes use of the temporal redundancy in one sequence, shown in Fig. 2.6. It is devoted to interceded frames because there exits some motion information. Temporal EC techniques are easier to implement and the complexity is less than some spatial domain algorithms. Temporal EC assumes the video data to be smooth or continuous as time goes on. The lost macroblock is constructed by front or rear frame with the same or near position.

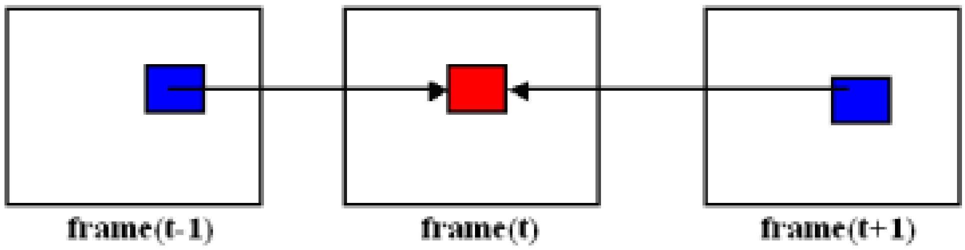


Fig. 2.6 Temporal domain EC

If only texture data is loss and the MV is still received, the lost macroblock can be decoded directly by MV with a little degradation, shown in Fig. 2.7.

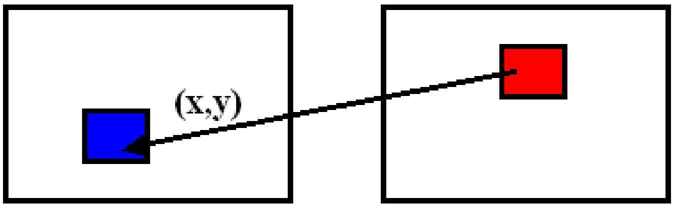


Fig. 2.7 The lost texture data

2.2.4 Conventional Temporal EC Techniques

An effortless method, temporal replacement (TR), replaced all damaged motion vectors by zero [18-19]. This methods works well for stationary areas but always collapsed for moving areas. The average, the weighted average or the median of the neighboring error-free motion vectors surrounding the damaged macroblock is utilize as the reconstructed motion vector. These methods also failed for the damaged macroblocks in areas with fast motion and on the object boundaries [3, 18-19].

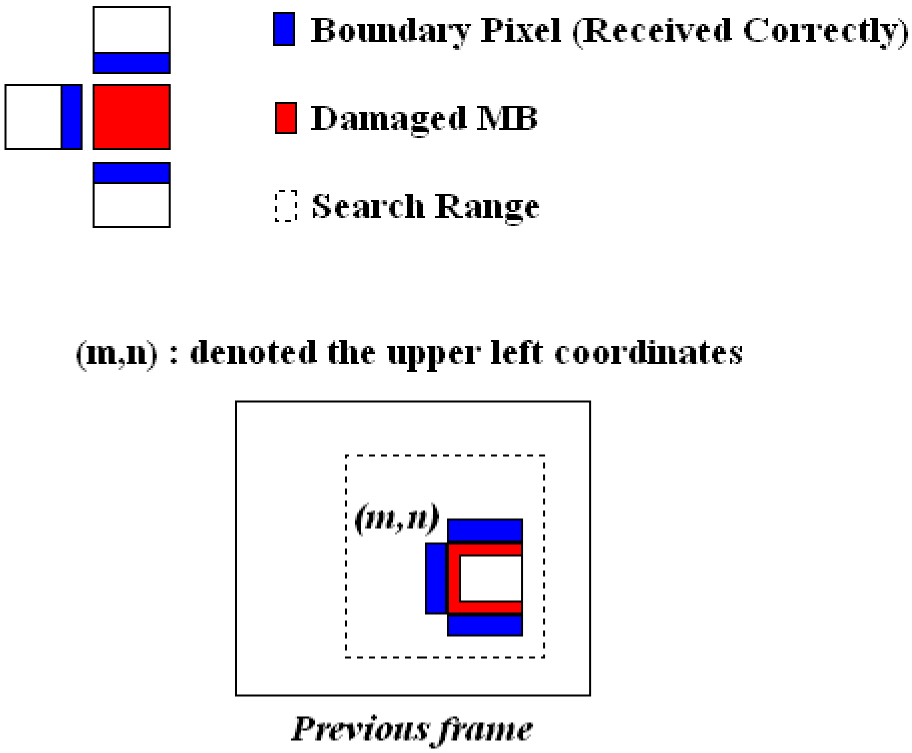


Fig. 2.8 Boundary matching algorithm

The boundary matching algorithm (BMA) [18] has been performed to identify the best replacement from a set of candidate motion vectors. The BMA search recovered motion vector technique always does well in minimizing the prediction error compared to reduce search techniques. This procedure selects a motion vector among a set of candidate vectors that minimizes the total variation between the boundaries of the damaged macroblock and those of the adjacent ones (the left, top and bottom macroblocks surrounding the damaged macroblock) as shown in Fig. 2.8. The method may fail for the damaged macroblocks in the areas with irregular motions or with low spatial correlation. Moreover, if the macroblocks surrounding a damaged macroblock are also lost, the BMA method may not perform well that was due to inappropriate motion vector estimation [20]. This situation is growing worse with increasing packet loss rate. In addition, slanting edges and rapid gray-level changes may cause an extensive variation and then the BMA method may be flunked. An improved BMA has been proposed to estimate the motion vector by further imposing smoothness properties on the diagonal and anti-diagonal directions for slanting edges and sub-pixel samples for rapid gray-level changes [20]. Feng et al. [21] also proposed a similar method to handle the cases of existing diagonal and anti-diagonal edges on the boundaries. Although the improved BMA approaches can cope with slating edges, they may fail in the case where there are edges across the damaged macroblock from left to right because of the lack of information on the left and right boundaries.

The motion field interpolation (MFI) method [21] was proposed to recover motion vector of the damaged macroblock. The damaged motion vector was approximated by using bilinear interpolation. The recovered motion vector is then used to conceal the damaged pixels. The MFI method provides a smoothly varying motion field that reduces blocking artifacts and compensates for more motion types, e.g. rotation and scaling. However, if control nodes are lost that will make ineffectual interpolating results.

2.3 Quality Evaluation

In general, the mean luminance PSNR (Peak Signal to Noise Ratio) is used to give a quantitative evaluation on the quality of the reconstructed frame in EC technique. We are also use the PSNR, which provides a first objective evaluation of the quality degradation for the considered video sequence. The PSNR parameter is defined as

$$PSNR = 10 \times \log \left[\frac{255^2}{\frac{1}{NM} \sum_{i=o}^{N-1} \sum_{j=o}^{M-1} \left(x(i,j) - x_R(i,j) \right)^2} \right], \qquad (2)$$

where $x(i,j)$ refers to the original image and $x_R(i,j)$ to the image after decoding; N and M define the image size.

Chapter 3

Error Concealment using Sub-Block Estimation

The conventional EC algorithms have been shown a poor performance for video sequences with scene changes or complex motion due to false coding mode estimation. A precise motion vector estimation method is essential for a robust temporal EC algorithm. In order to perform precise motion vector estimation, a damaged macroblock was further divided into 4 non-overlapping sub-blocks (S_1, S_2, S_3, and S_4), as showing in Fig. 3.1.

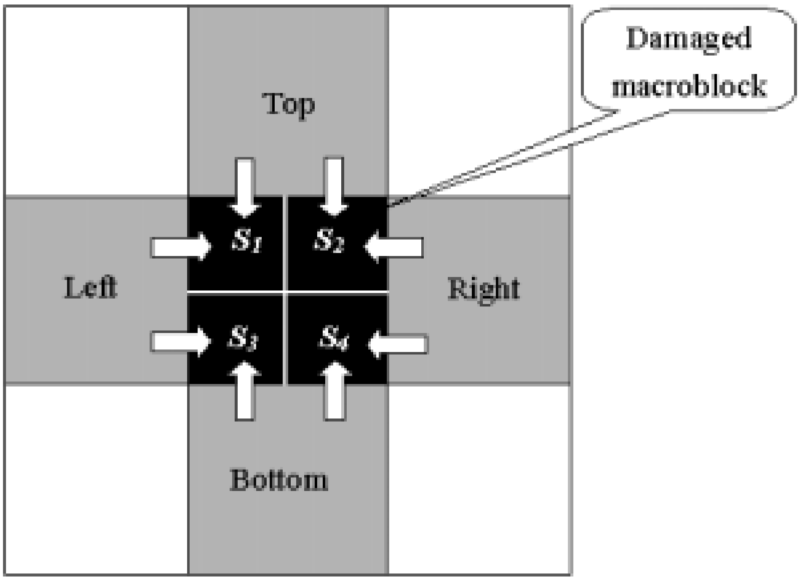


Fig. 3.1 A damaged macroblock (black part) was divided into four sub-blocks. The motion vector for each sub-block was estimated by using the information of the nearest error-free macroblocks.

3.1 Temporal Sub-Block EC Algorithm

The proposed temporal sub-block EC (TSEC) algorithm estimates the motion vectors of the sub-blocks by exploiting the information of four adjacent macroblocks (top, bottom, left, and right macroblocks) surrounding the damaged macroblock. Each motion vector of sub-block is estimated by using the information of the nearest adjacent macroblocks. Let MV_X represents motion vector of the damaged macroblock X. MV_T, MV_L, MV_R and MV_B represent the motion vector of top, left, right, and bottom macroblock surrounding X, respectively. The motion vector for S_1, S_2, S_3, and S_4 are estimated as follows.

$$MV_{S1} = \frac{MV_T + MV_L}{2}, \text{ if both } MV_T \text{ and } MV_L \text{ are undamaged;}$$

$$MV_{S2} = \frac{MV_T + MV_R}{2}, \text{ if both } MV_T \text{ and } MV_R \text{ are undamaged;}$$

$$MV_{S3} = \frac{MV_B + MV_L}{2}, \text{ if both } MV_B \text{ and } MV_L \text{ are undamaged;}$$

$$MV_{S4} = \frac{MV_B + MV_R}{2}, \text{ if both } MV_B \text{ and } MV_R \text{ are undamaged.} \quad (3)$$

If one of the neighboring macroblocks was also lost, the motion vector of the sub-block will be replaced as the undamaged motion vector. Unfortunately, if no valid motion vector of the neighbors is available, the simple temporal replacement with a zero motion vector is used for the sub-block.

Moreover, the proposed TSEC method performs an area search within the previous frame for the best boundary match surrounding a damaged macroblock. The area search aims to achieve the more accurate motion vector estimation. The search area is always in a small range. In our experiments, the search area is selected as *+/-1* pixel width in both horizontal and vertical directions. Similar to the BMA, the proposed TSEC method selects a motion vector among a set of candidate vectors that minimizes the total variation between the boundaries of the damaged macroblock in the search area. We also found that when the macroblock in the border of a frame was damaged, the motion vectors for sub-blocks may result an inaccurate estimation. The reason is that the estimation procedure lacked enough information for making a good prediction of motion vector. Thus the BMA method is utilized to conceal the border damaged macroblocks that is shown in Fig. 3.2.

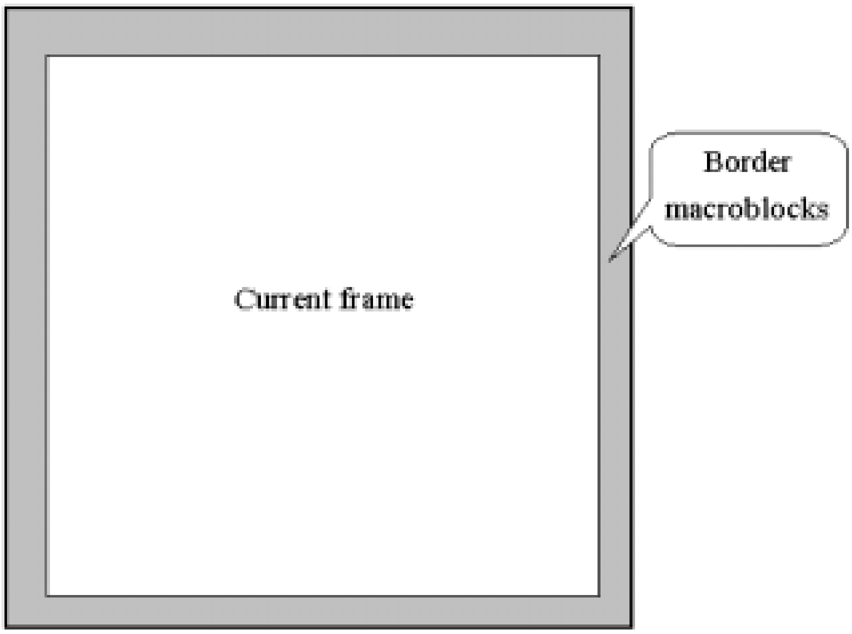


Fig. 3.2 Border macroblocks (gray part) in the current frame

3.2 Experimental Results

Three image sequences, “foreman” (300 frames, CIF), “table tennis” (150 frames, CCIR601) and “flower garden” (150 frames, CCIR601) were used in the simulations. The image sequences were encoded by MPEG-2 standard. Each group of picture (GOP) in encoded sequenced consists of 18 frames (1 I-frame, 5 P-frames, 12 B-frames). The size of macroblock is 16×16. Transmission errors are simulated by using a two-state Markov model, namely the Gilbert channel model. The comparisons for EC were made in average macroblock missing rates 1%, 5%, 10% and 20%. The peak signal-to-noise ratio (PSNR) is used to give a quantitative evaluation on the quality of the reconstructed frame.

Table 3.1 PSNR mean values comparison for three test sequences with different EC methods.

Image sequence	EC method	Average macroblock missing rate			
		1%	5%	10%	20%
Foreman	TR	37.81	30.96	27.53	24.71
	BMA	41.54	34.76	30.99	28.20
	TSEC	42.22	35.24	31.40	28.54
Table tennis	TR	40.92	30.39	27.71	23.86
	BMA	42.15	32.03	28.87	25.75
	TSEC	43.24	32.98	29.58	26.21
Flower garden	TR	28.34	21.18	18.18	15.56
	BMA	32.45	24.40	21.58	18.75
	TSEC	33.99	25.71	22.86	19.81

The proposed TSEC method and other popular existing temporal EC approaches are implemented in this study. We compare three temporal EC methods, the TR, the BMA, and our new TSEC in the simulations. In this paper, the size of sub-block is 8 × 8 and the search area is selected as +/-1 pixel width. Table 3.1 shows the average PSNR values for reconstructed image sequences by using the three EC methods in average packet loss rates 1%, 5%, 10% and 20%. Fig. 3.3, Fig. 3.4 and Fig. 3.5 show the PSNR values for individual frame of the reconstructed test sequences “foreman”, “table tennis” and “flower garden”, respectively. The proposed TSEC method achieves the better performance than other methods, especially for “flower garden” image sequence.

Fig. 3.6(a) shows an error-free frame in test sequence “flower garden” and Fig. 3.6(b) is the damaged frame with no-overlapped stripes lost. The damaged macroblocks in the frame are replaced with gray-level 0. The reconstructed frames using the TR, BMA and proposed TSEC methods are shown in Fig. 3.6(c), Fig. 3.6(d) and Fig. 3.6(e), respectively. It was apparent that the reconstructed frame produced by TR was inferior to that of the BMA and the TSEC. To show the differences of the reconstructed frame using the BMA and the TSEC method, the enlarged portions in the reconstructed frames are given in Fig. 3.6(f). We observe that the proposed TSEC algorithm can really diminish the occurrence of blocking effects in reconstructed frame. Similarly, Fig. 3.7(a) and Fig. 3.7(b) show the undamaged and damaged frames from test sequence “table tennis”, respectively. The reconstructed frames using the TR, BMA and TSEC methods are shown in Fig. 3.7(c), Fig. 3.7(d) and Fig. 3.7(e), respectively. Fig. 3.7(f) shows the enlarged portions in the reconstructed frames of Fig. 3.7(d) and Fig. 3.7(e). The BMA shows that can produce acceptable recovered frames. However, it cannot successfully reconstruct the macroblocks that contain edges across the left and right boundaries. That is because the information of left and right macroblocks surrounding the damaged macroblock is not available. The simulation results demonstrate that the proposed TSEC method can exploit the interior information of the predicted macroblock to accurately estimate the lost motion vectors.

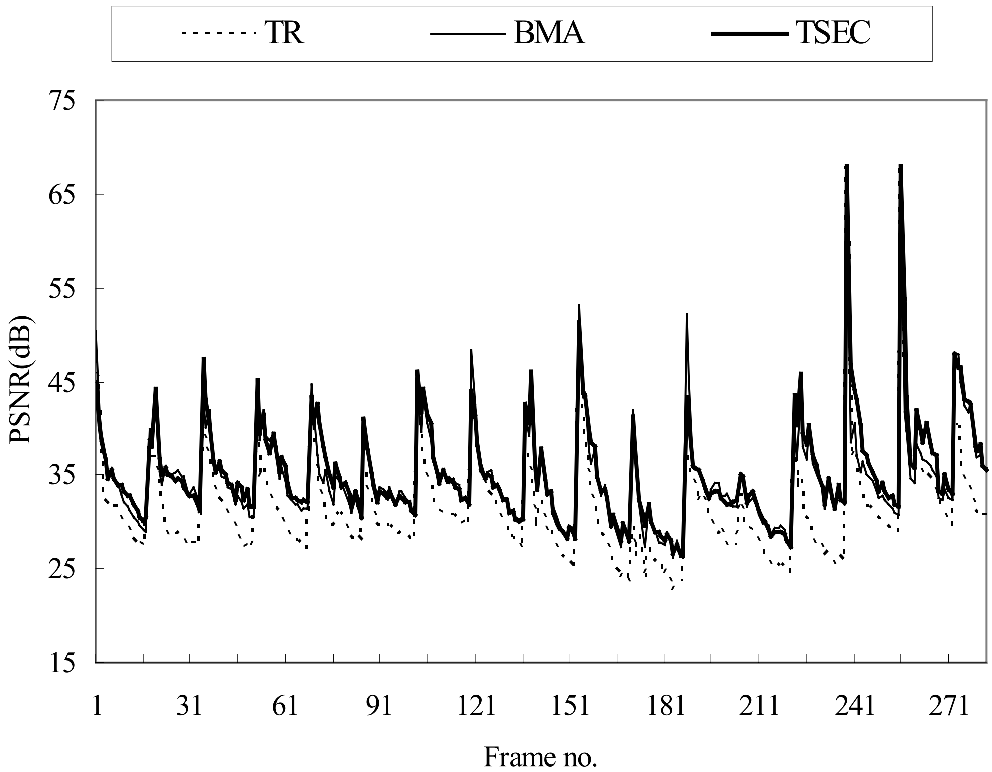


Fig. 3.3 PSNR values for frames of the test sequence "foreman." The sequence is affected by data losses with a 5% of macroblock missing.

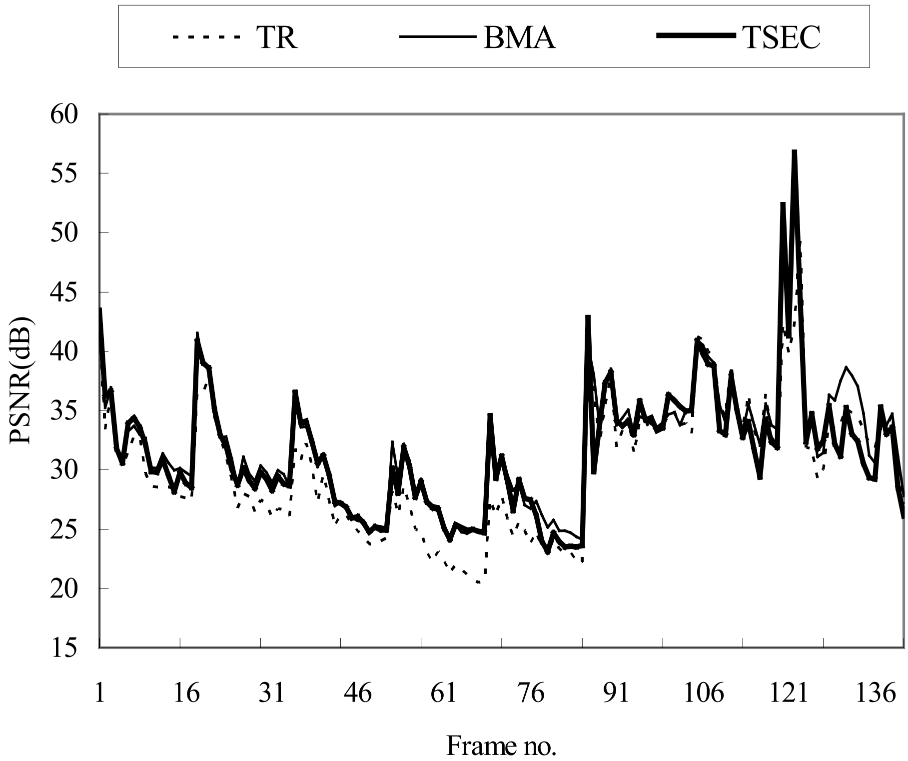


Fig. 3.4 PSNR values for frames of the test sequence "table tennis." The sequence is affected by data losses with a 5% of macroblock missing.

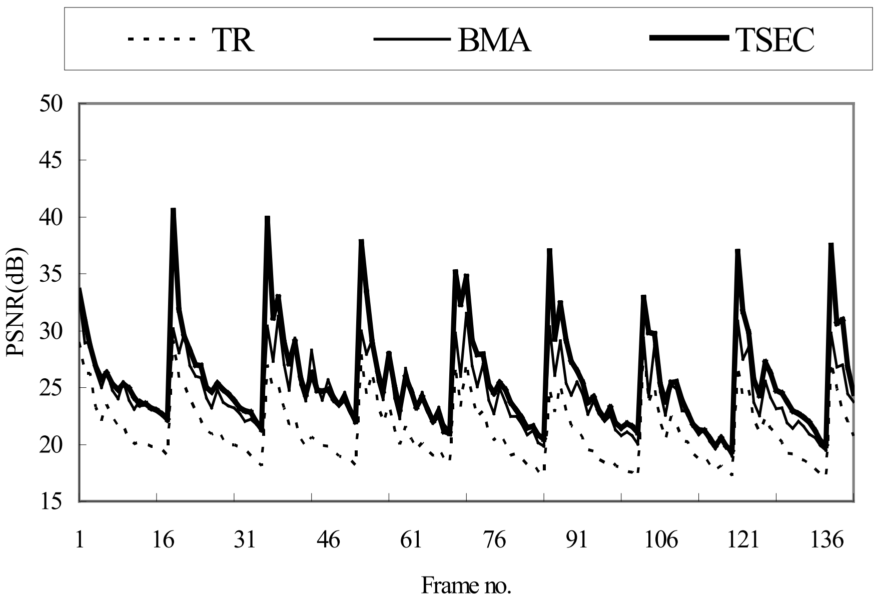


Fig. 3.5 PSNR values for frames of the test sequence "flower garden." The sequence is affected by data losses with a 5% of macroblock missing.

(a)

(b)

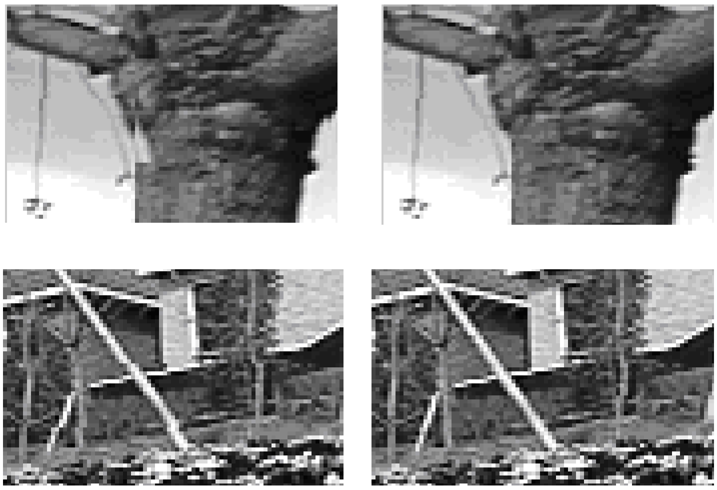
Fig. 3.6 Subjective video quality comparison for frame 37 in test sequence “Flower Garden”: (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; (e) reconstructed frame using the proposed TSEC method; and (f) magnified portion of (d) and (e), respectively. (continued)

(c)

(d)

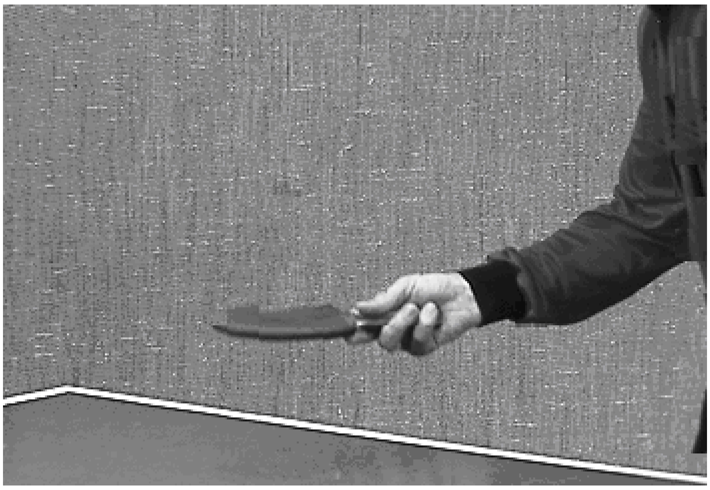
Fig. 3.6 Subjective video quality comparison for frame 37 in test sequence "Flower Garden": (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; (e) reconstructed frame using the proposed TSEC method; and (f) magnified portion of (d) and (e), respectively. (continued)

(e)

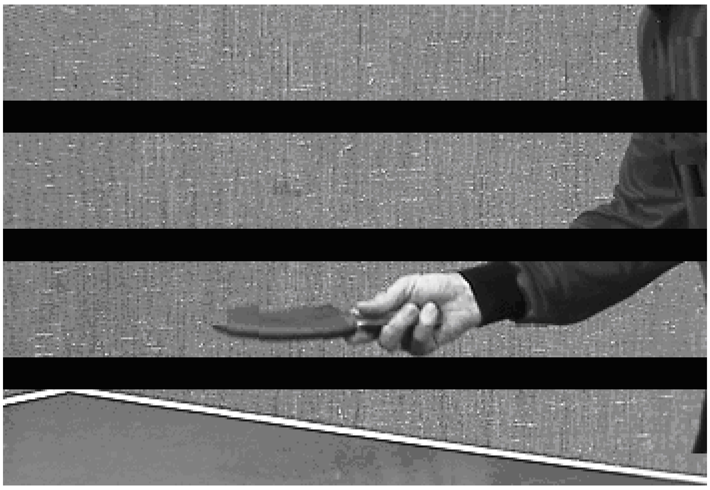


(f)

Fig. 3.6 Subjective video quality comparison for frame 37 in test sequence “Flower Garden”: (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; (e) reconstructed frame using the proposed TSEC method; and (f) magnified portion of (d) and (e), respectively.

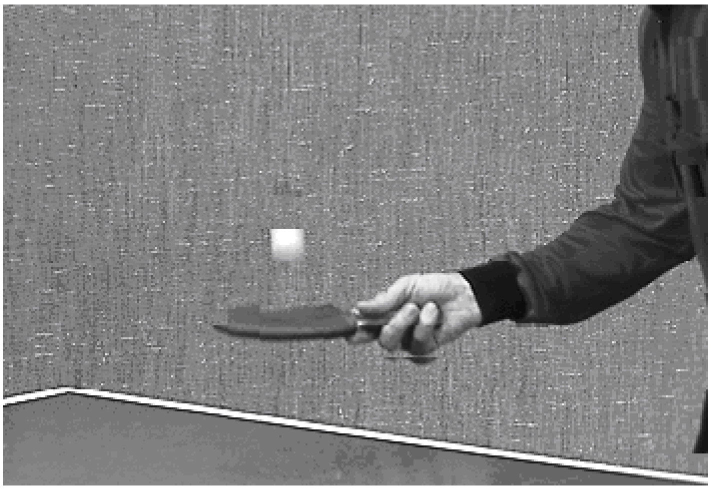


(a)

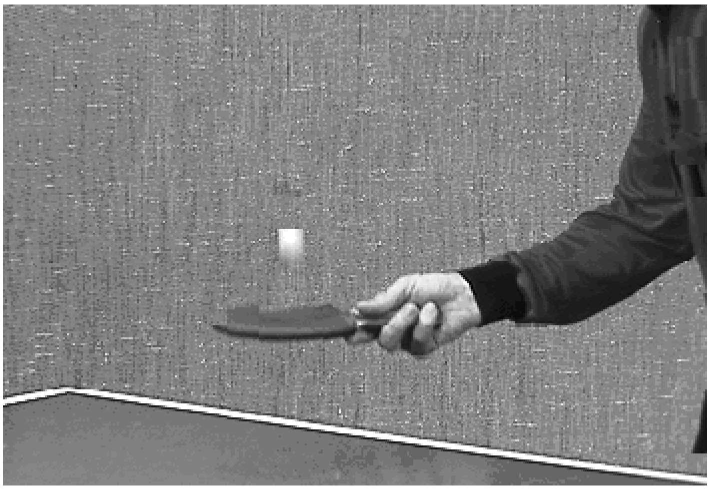


(b)

Fig. 3.7 Subjective video quality comparison for frame 26 in test sequence “Table Tennis”: (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; (e) reconstructed frame using the proposed TSEC method; and (f) magnified portion of (d) and (e), respectively. (continued)

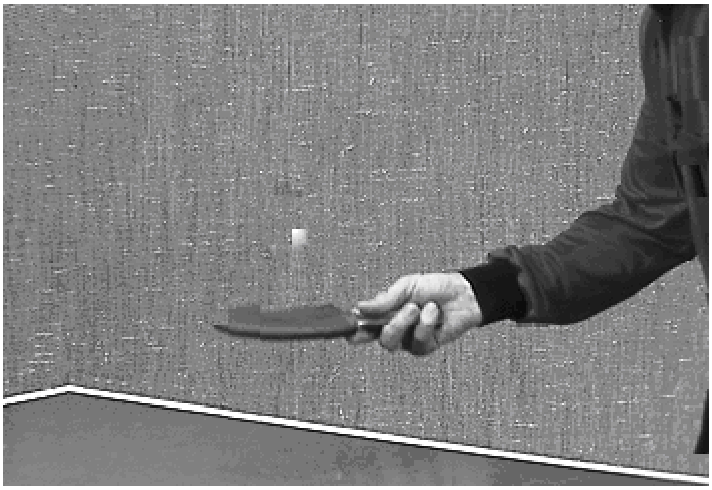


(c)

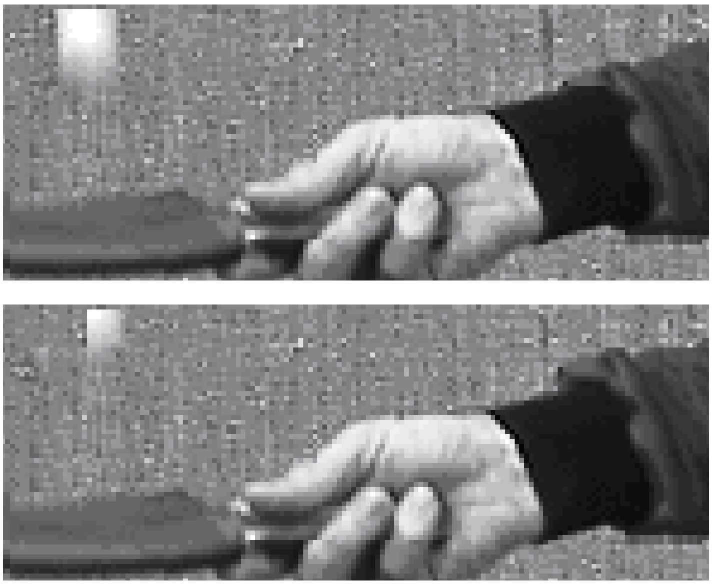


(d)

Fig. 3.7 Subjective video quality comparison for frame 26 in test sequence "Table Tennis": (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; (e) reconstructed frame using the proposed TSEC method; and (f) magnified portion of (d) and (e), respectively. (continued)



(e)



(f)

Fig. 3.7 Subjective video quality comparison for frame 26 in test sequence “Table Tennis”: (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; (e) reconstructed frame using the proposed TSEC method; and (f) magnified portion of (d) and (e), respectively.

Chapter 4

Error Concealment using Self-Organizing Map

The SOM has been shown as an effective neural network model for the visualization of high-dimensional data. This neural model has the ability to e SOM learning algorithm convert data items from a high-dimensional space onto simple reference vectors in a low-dimensional space. Thus this study employs the SOM as a predictor to estimate the motion vector of the damaged macroblock more accurately.

4.1 Motion Vector Estimation Using SOM

An SOM neural network [5] contains an input layer, a single hidden layer, and a mapping array of output. It refers to the ability of unsupervised learning. The number of input neurons ensures from the dimension of the input vectors. In general, the SOM model defines a mapping from the higher dimension of input data space onto a regular two-dimensional mapping array is shown in Fig. 4.1. With every neuron in the mapping array, a parametric weight vector produced by learning algorithm is associated. An input vector will compare with all weight vectors, and the best match is defined as the SOM response. For a more useful analysis, each neuron in the mapping array may also be marked a class label using training samples. The learning algorithm for SOM is shortly described as the following steps:

1. Initialization of the SOM weights to small random values. Establish the initial radials of the region for weight modification.
2. Presentation of an N-dimension input vector $\boldsymbol{x}$.
3. Calculation of the distance d_j between the current input vector and the weight vectors of all mapping array neurons according to the Euclidean

distance function

$$d_j = \sum_{i=1}^{N} \left(x_i - w_{ji} \right)^2 \tag{4}$$

where x_i is the i-th component of the input vector and w_{ij} is the i-th component of the weight vector $\boldsymbol{w}_j$.

4. Determination of the output neuron $\boldsymbol{o}$ which is the one with minimal distance

5. Update of all weight vectors of neurons lying with in the region of weight modification. Weight modification is given by

$$\Delta w_{ji} = o_j \eta \left(x_i - w_{ji} \right) \qquad (5$$

where η is the learning rate parameter.

6. Repeat the steps above by presenting a new input vector until the given stop criterion is achieved.

During the weight modification phase of the learning algorithm, the learning rate parameter and radial of the region of weight modification could be monotonically decreasing over time. This work can be used to minimize the distortion of the SOM. The SOM models perform high unsupervised learning capability and computational efficiency. The proposed temporal SOM EC system estimates the motion vectors of damaged macroblocks by adding the final SOM mapping array are shown in Fig. 4.2.

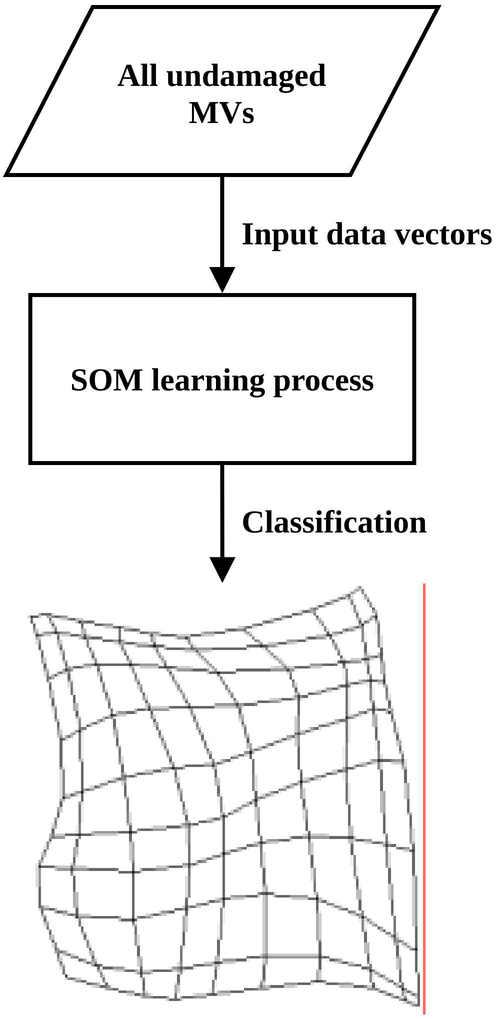


Fig. 4.1. Shows how this works with a two-dimensional vector.

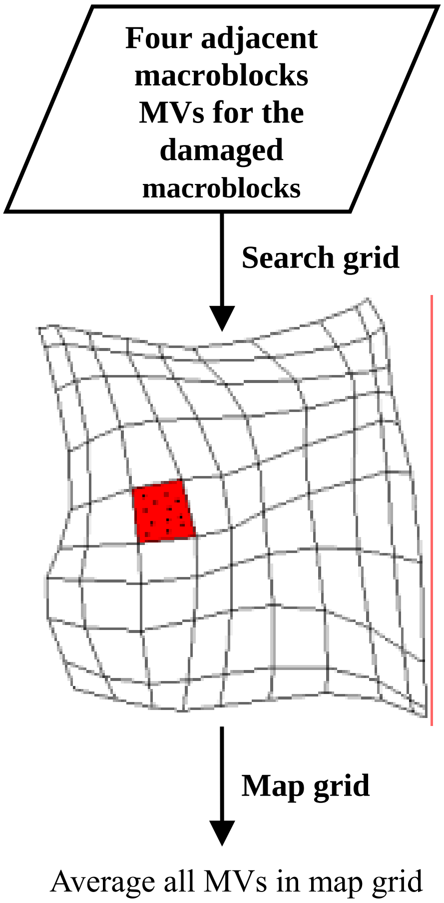


Fig. 4.2. Motion vector estimation using SOM

4.2 Adaptive EC Algorithm using SOM

Before any EC techniques can be applied to the compressed images, the locations of damaged macroblocks are necessary first to be found out. In this paper, we only focus on the problem of concealing the error macroblocks for macroblock-based image coding systems. Thus we assume that the locations of damaged macroblocks are known and discuss techniques for concealing the detected errors. The proposed SOM temporal EC algorithm estimates the motion vectors of the damaged macroblocks by exploiting the information of four adjacent macroblocks (top, bottom, left, and right macroblocks). Let MV_X represents motion vector of the damaged macroblock ***X***. MV_T, MV_L, MV_R and MV_B represent the motion vector of top, left, right, and bottom macroblock surrounding ***X***, respectively. By the spatial redundancy between the motion vectors of adjacent macroblocks, the SOM model utilizes the motion vectors of macroblocks surrounding the damaged macroblock to estimate the motion vector of the damaged macroblock is shown in Table. 4.1.

Table 3.1 Input data format as SOM tracing source

MV_T		MV_L		MV_R		MV_B	
-9	14	-2	0	-2	0	-2	-1
-9	14	-5	6	-2	0	-2	0
—	—	-2	0	—	—	-2	0
0	0	—	—	-2	0	-2	0
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
-2	-3	-5	1	-2	0	-2	0

The proposed algorithm described previously assumes that the errors are localized to separated macroblocks in the decompressed frames. In fact, errors typically propagate through several consecutive macroblocks for compressed frames and then yield some of adjacent lost macroblocks. In this condition, the conventional temporal EC techniques would fail to accurately estimate the motion vector of the damaged macroblock because there is no enough information. On the contrary, the SOM model has the possibility to estimate the motion vector of damaged macroblock by using incomplete input motion vectors. The SOM will compute the distance calculations and reference vector modification steps using the available data components. For example, partial data can still be used to determine the distribution statistics of the available vector components [5, 22].

Moreover, the proposed SOM method performs a search within the previous frame for the best boundary match surrounding a damaged macroblock. Similar to the BMA, the proposed SOM method selects a motion vector among a set of candidate vectors that minimizes the total variation between the boundaries of the damaged macroblock. Thus candidate vectors of the BMA method are added to the proposed SOM EC to conceal the damaged macroblocks. Clearly, the proposed temporal SOM EC method is always practical even if the motions in a frame are complex or irregular. The whole proposed method utilizes only one manner of neural network model - the SOM network. Thus, the architecture of the proposed temporal SOM EC algorithm is simple, redressing easily, and suitable for hardware design. Moreover, the SOM neural networks are highly parallel computer architecture and, thus, offer the potential for real-time applications.

4.3 Experimental Results

In this study, three image sequences, “bus” (300 frames, CCIR601), “football” (150 frames, CCIR601) and “flower garden” (150 frames, CCIR601) were used in the simulations. The image sequences were encoded by MPEG-2 standard. Each group of picture (GOP) in encoded sequenced consists of 18 frames (1 I-frame, 5 P-frames, 12 B-frames) and the size of macroblock is 16×16. We perform a two-state Markov model, namely the Gilbert channel model, to simulate the transmission errors. The comparisons for EC methods were made in average macroblock missing rates 1%, 5%, 10% and 20%.

Table 4.2 Average PSNR values for the test sequences with different EC methods.

Image sequence	EC method	Average macroblock missing rate			
		1%	5%	10%	20%
Bus	TR	26.46	26.46	23.37	20.95
	BMA	29.86	29.86	26.13	23.63
	SOM	30.66	31.33	27.40	24.98
Football	TR	34.37	25.56	23.07	20.22
	BMA	35.80	27.04	24.35	21.52
	SOM	37.69	28.98	26.27	23.31
Flower garden	TR	28.22	21.15	18.14	15.56
	BMA	33.40	24.72	21.38	18.75
	SOM	35.40	26.10	22.66	20.15

The proposed SOM method and other popular existing temporal EC approaches are implemented in this study. In order to prove the precision of the proposed SOM estimation, we compare the results using the proposed SOM EC algorithm and simulation results of TR method that replaced all damaged motion vectors by zero and the BMA method proposed in [19]. Table 4.2 shows the average PSNR values for reconstructed image sequences by using the three EC methods in average packet loss rates 1%, 5%, 10% and 20%. The proposed SOM method achieves the better performance than other methods. Fig. 4.3, Fig. 4.4 and Fig. 4.5 show the PSNR values for individual frame of the reconstructed test sequences “bus”, “football” and “flower garden”, respectively.

Fig. 4.6(a) shows an error-free frame in test sequence “flower garden” and Fig. 4.6(b) is the damaged frame with no-overlapped stripes lost. The damaged macroblocks in the frame are replaced with gray-level 0. The reconstructed frames using the TR, BMA and the proposed SOM method are shown in Fig. 4.6(c), (d) and (e), respectively. It was apparent that the reconstructed frame produced by TR was inferior to that of the BMA and the proposed SOM method. The recovered image using the proposed SOM method shows the improved subjective quality than other recovered images. Similarly, Figs. 4.7(a) and (b) show the undamaged and damaged frames from test sequence “bus”, respectively. The reconstructed frames using the TR, BMA and the proposed SOM method are shown in Figs. 4.7(c), (d) and (e), respectively. The BMA shows that can produce acceptable recovered frames. However, it cannot successfully reconstruct the macroblocks that contain edges across the left and right boundaries. That is because the information of left and right macroblocks surrounding the damaged macroblock is not available. We observe that both the smooth and detailed regions in the reconstructed frames can obtain the good visual quality using the proposed SOM EC algorithm. From the simulation results, we find that the proposed SOM algorithm has very good performance.

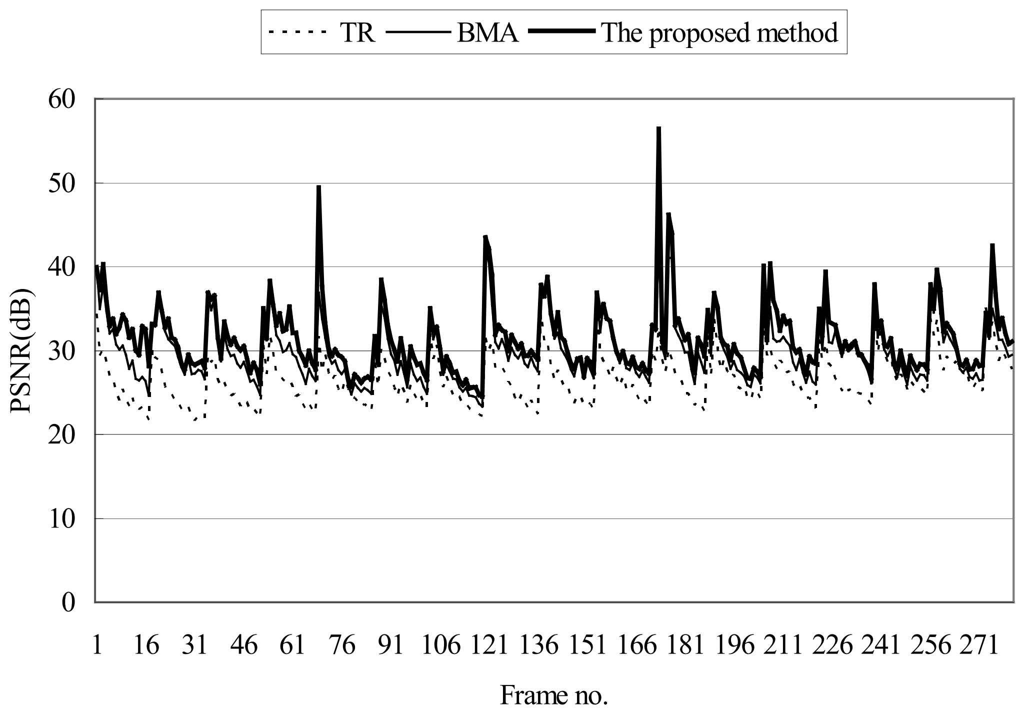


Fig. 4.3 PSNR values for frames of the test sequence “bus.” The sequence is affected by data losses with a 5% of macroblock missing.

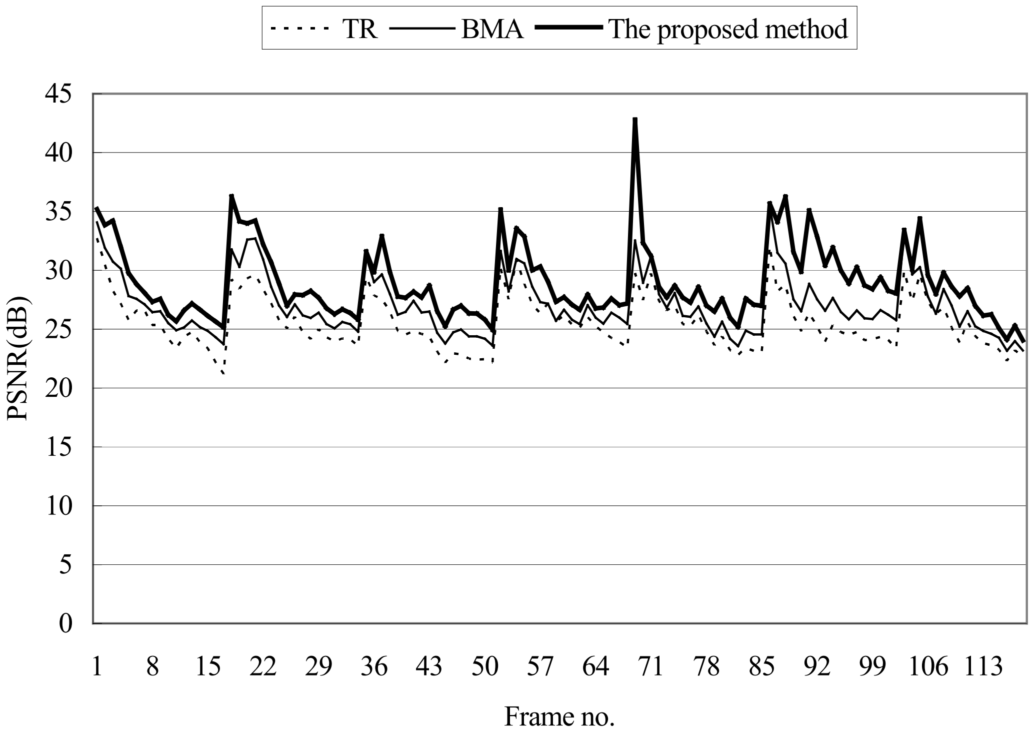


Fig. 4.4 PSNR values for frames of the test sequence “football.” The sequence is affected by data losses with a 5% of macroblock missing.

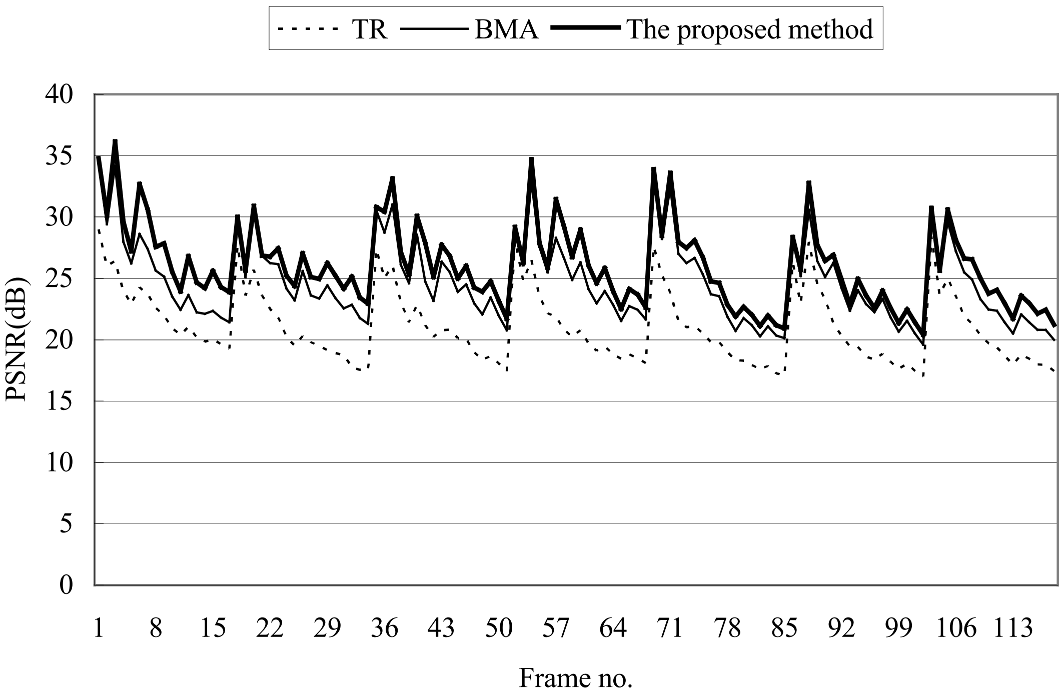


Fig. 4.5 PSNR values for frames of the test sequence “flower garden.” The sequence is affected by data losses with a 5% of macroblock missing.

To quantify the computational complexity of the EC methods, we measured the execution time required for each technique. The average execution time of reconstructed frames using the TR, the BMA and the proposed SOM method are 0.062, 3.4138 and 3.8578 second, respectively. In the proposed SOM method, the average execution time is slightly larger than other two methods. However, the proposed SOM method achieves the better video quality. We conclude that the proposed SOM method robust EC technique can be a useful alternative to existing techniques due to its low computational complexity and acceptable PSNR performance.

(a)

(b)

(c)

Fig. 4.6 Subjective video quality comparison for frame 39 in test sequence "Flower Garden": (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; and (e) reconstructed frame using the proposed SOM method. (continued)

(d)

(e)

Fig. 4.6 Subjective video quality comparison for frame 39 in test sequence “Flower Garden”: (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; and (e) reconstructed frame using the proposed SOM method.

(a)

(b)

(c)

Fig. 4.7 Subjective video quality comparison for frame 73 in test sequence "Bus": (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; and (e) reconstructed frame using the proposed SOM method. (continued)

(d)

(e)

Fig. 4.7 Subjective video quality comparison for frame 73 in test sequence “Bus”: (a) original frame; (b) corrupted frame; (c) reconstructed frame using the TR method; (d) reconstructed frame using the BMA; and (e) reconstructed frame using the proposed SOM method.

Chapter 5

Conclusions

This thesis presents two novel temporal EC techniques, aimed at masking the loss of visual information due to erroneous transmission of coded digital video over unreliable networks. In the first method, the proposed TSEC method aims at masking the loss of visual information due to erroneous transmission of coded digital video over unreliable networks. We utilized the motion vector estimation for sub-blocks in the damage macroblock to overcome the disadvantage of the BMA. Experimental results demonstrated that the proposed TSEC approach is able to significantly reduce the artifacts that characterize by conventional temporal EC techniques and increasing both the subjective and objective quality of the reconstructed frames. By comparing with other methods, the proposed TSEC method obtains the superior performance for EC on image sequences with variant motion. From the experimental results, we find that the proposed TSEC method is expected to be a useful EC algorithm for MPEG video coding systems. However, this approach entails a considerable amount of processing complexity at the decoder. The computational complexity needs to decrease in the future. We also desire that the proposed TSEC method can extend to other video compression applications, such as videoconferencing, video on demand, and high-definition TV.

The second method using the SOM neural network model overcomes the disadvantage of boundary matching algorithm at fail in areas with unsmooth motion and also for areas with low spatial correlation. The proposed SOM method is able to reconstruct motion vector of the damaged macroblock by using incomplete information and increasing both the subjective and objective quality of the reconstructed frames even under the high packet losses. Even though the motion is extensive, the proposed SOM method increase decoded video quality by using the nonlinear statistical relationships feature of the motion. The proposed SOM method can estimate the motion vectors of the corrupted macroblocks even in the sequences including fast moving objects. Thus the high-motion sequences have better performance than the small-motion sequences. The proposed SOM temporal EC will be applicable to the video decoder requiring low computation complexity and the robust EC to high scene activity.

Reference

[1] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall, *MPEG Video Compression Standard*. London, U.K.: Chapman & Hall, 1997.

[2] Y. Wang and Q. F. Zhu, "Error control and concealment for video communication: a review," *Proc. IEEE*, Vol. 86, May 1998, pp. 974-997.

[3] J. W. Suh and Y. S. Ho, "Error concealment based on directional interpolation," *IEEE Trans. on Consumer Electronics*, Vol. 43, Aug. 1997, pp. 295-302.

[4] L. Atzori, F. G. B. De Natale and C. Perra, "A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping (BMA-MBW)," *IEEE Trans. on Multimedia*, Vol. 3, Sept. 2001, pp. 326–338.

[5] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Heidelberg, 1995.

[6] N. V. Swindale1 and H. U. Bauer, "Application of Kohonen's self-organizing feature map algorithm to cortical maps of orientation and direction preference," *The Royal Society*, Jan. 1998, pp. 827-838.

[7] D. Cook, A. Buja, J. Cabrera, and C. Hurley, "Grand tour and projection pursuit," *Journal of Computational and Graphical Statistics*, Vol. 4, no. 3, 1995, pp. 155-172.

[8] M. A. Kraaijveld, J. Mao, and A. K. Jain, "A nonlinear projection method based on Kohonen's topology preserving maps," *IEEE Transactions on Neural Networks*, Vol. 6, May 1995, pp. 548-559.

[9] J. Vesanto, "SOM-based data visualization methods," *Intelligent Data Analysis*, Vol. 3, April 1999, pp. 111-126.

[10] I. S. Dhillon and D. S. Modha, "Concept decompositions for large sparse text data using clustering," *Machine Learning*, Vol. 42, Jan. 2001, pp. 143-175.

[11] W. K. Pratt, *Digital Image Processing*, Wiley-Inter-science Publication, 1978.

[12] H. Sun, K. Challapali and J. Zdepski, “Error concealment in digital simulcast AD-HDTV decoder,” *IEEE Trans. on Consumer Electronics*, Vol. 38, No. 3, Aug. 1992, pp. 108-116.

[13] Q. F. Zhu, Y. Wang, and L. Shaw, “Coding and cell-loss recovery in DCT-based packet video,” *IEEE Trans. on Circuits and Systems for video Technology*, Vol. 3, Jun. 1993, pp. 248-258.

[14] M. H. Jo and W. J. Song, “Error concealment for MPEG-2 video decoders with enhanced coding mode estimation,” *IEEE Trans. on Consumer Electronics*, Vol. 46, Nov. 2000, pp. 962-969.

[15] M. B. Ahmad, D. Y. Kim, K. S. Roh, T. S. Choi, “Motion vector estimation using edge oriented block matching algorithm for video sequences,” in *Proc. International Conference on Image Processing*, Vol. 1, Sep. 2000, pp. 860–863.

[16] Y. C. Lee, Y. Altunbasak, R. Mersereau, “A temporal error concealment method for MPEG coded video using a multi-frame boundary matching algorithm,” in *Proc. International Conference on Image Processing*, Vol. 1, Oct. 2001, pp. 990–993.

[17] S. Cen and P. C. Cosman, “Decision trees for error concealment in video decoding,” *IEEE Trans. on Multimedia*, Vol. 5, March 2003, pp. 1–7.

[18] W. M. Lam, A. R. Reibman, and B. Liu, “Recovery of lost or erroneously received motion vectors,” in *Proc. IEEE ICASSP*, Vol. 5, Mar. 1993, pp. 417–420.

[19] J. W. Suh and Y. S. Ho, “Recovery of motion vectors for EC,” in *Proc. IEEE Region 10 Annual Conference*, Sep. 1999, pp. 750-753.

[20] M. J. Chen, C. S. Chen, M. –C. Chi, "Recursive block-matching principle for error concealment algorithm," in *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS2003)*, Vol. 2, May 2003, pp. 528-531.

[21] M. Al-Mualla, N. C. Canagarajah, and D. R. Bull, "Temporal error concealment using motion field interpolation," *IEE Electron. Letter*, Vol. 35, No. 3, Feb. 1999, pp. 215-217.

[22] T. Samad, S. A. Harp, "Self-organization with partial data," *Network: Computation in Neural Systems*, 1992, pp. 205-212.