Content-Adaptive Spatial Error Concealment for Video Communication

Zhang Rongfu, Zhou Yuanhua and Huang Xiaodong

Abstract — *When bit errors occur during transmission and cannot be corrected by an error correction scheme, error concealment is needed to mask the damaged image at receiver. In this paper, a content-adaptive spatial error concealment algorithm is presented. By using edge information extracted from the surrounding blocks, the proposed algorithm firstly classifies each error block (EB) into one of three categories with different contents. And then the EB is reconstructed by appropriate methods to the category of it belonging to. Experimental results demonstrate that the proposed algorithm obtain good subjective quality whether the content in the EB is with high or low details.*

Index Terms —Error concealment, Spatial interpolation, Block matching, Video communication.

I. INTRODUCTION

Video communications are becoming the most important source of traffic in modern data networks. Image and video sequences are used in a wide range of applications from home entertainment to complex industrial systems. Due to the high demand in terms of bandwidth, most of current image/video coding standards, such as JPEG, H.26x and MPEG, adopt discrete cosine transform (DCT), motion compensation (MC), and variable length coding (VLC) techniques. These techniques obtain high compression ratio. However, any data corruption, independently of the cause and of the extent, may arise in serious error propagation problems in the decoding process. Therefore, when video signals are compressed and transmitted over unreliable channels, some strategy must be employed to make the quality of the decoded image acceptable.

Among several approaches for above problem, the error concealment (EC) technique [1] is widely investigated for its possibility of general application. The EC scheme attempts to recover the error blocks (EBs) by utilizing information from spatially or temporally adjacent blocks. The temporal schemes conceal the EBs with small motion, by exploiting the data in the adjacent frames. However, to the damaged blocks with large motion or error in the first intracoded frame (I frame) of a scene, spatial schemes are preferred. The algorithms proposed in [3], [4], [5] and [6] interpolate pixel values in EB by using pixels in correctly reconstructed blocks neighboring to the EB. These methods work well only on the smooth or regular blocks. However, for the irregular or high-detail blocks, their performance deteriorates followed by noticeable blurring image blocks and the reconstructed block does not match well with the surrounding blocks. Neighborhood matching EC (NMEC) [7] is proposed by making use of a spatial kind of a priori information—blockwise similarity within the image. An algorithm, called the spatial split-match EC method, is introduced in [8], which attempts to spatially match top and bottom neighboring regions and conceals the region in the direction of the match. Although the methods used in [7] and [8] can successfully conceal EBs with high details, they will produce the undesired blocky artifacts and edge discontinuity when the content of local image is with low details. In addition, the latter two methods have the disadvantage of high computation complexity for searching a suitable block.

In order to improve concealed image quality and speed up the concealment, a content-adaptive error concealment algorithm is proposed in this paper. The first step of the algorithm is to estimate the content of the EB and classify the EB into one of the three categories: uniform block, edge block and texture block. And the second step is to conceal the EB by a suitable method according to the category of the EB belonging to.

The rest of this paper is organized as follows. In the next section, an overview of the proposed EC algorithm is presented. The two modules of the algorithm are described in details in section III and IV respectively. Section V reports the experimental results in terms of perceptive quality of the concealed image/video. Finally the conclusions are drawn in section VI.

II. OVERVIEW OF THE PROPOSED EC ALGORITHM

A block diagram of realtime video communications system is shown in Fig. 1. The input video is encoded using an appropriate video compression syntax. The transport coder in the figure is used to convert the bit-stream output from the source coder into data units suitable for transmission. The

[1] This work was supported in part by the High Technology Research and Development Program of China under Grant No. 2003AA103810 and Fund of Shanghai Education Committee under Grant No. 03GQ22.

The authors are with Institute of Image Communication and Information Processing, Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030, P.R.China. (email: zrfxm@163.com)

Contributed Paper
Manuscript received December 17, 2003

channel is transmission medium, e.g. Broadcast, ATM, Internet, and so on. At the receiver side, the inverse operations are performed to obtain the reconstructed video signal for display. If the channel is error prone, bit error or packet/cell missing may occur and result in lost of block data. Therefore, it is necessary to conceal the error when decoding.

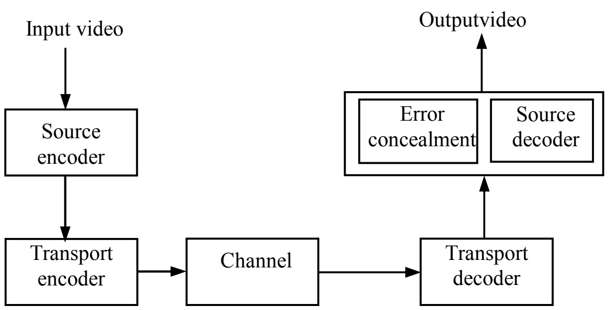


Fig.1 Video communication system diagram

The framework of the proposed EC algorithm is shown in Fig.2. The coded video bitstream is detected to determine if there is error block and signs the position of the error block if EB exits. In order to use the redundancy in all the available surrounding blocks, the concealment function is called untill all the valid blocks in the frame have been completely decoded and the decoded frame without being concealed is stored in a frame store memory. And then all EBs in the current frame are restored by the EC algorithm proposed in this paper. The damaged frame is reconstructed by replacing all those EBs with corresponding restored ones. As shown in the right part of Fig.2, the proposed EC algorithm consists of two main modules, i.e., the block classification module and concealment module. The detailed design and implementation of these two modules will be described in the following two sections.

III. CONTENT ESTIMATION AND BLOCK CLASSIFICATION

As introduced in the first section, one EC method may perform better than the others depending on the content of the video sequence. Therefore, it is necessary to select a suitable EC method to conceal the EB based on the local edge characteristics. However, the real values of pixels in the EB are not available when conceals the EB. In order to conceal each EB with a suitable EC method, the content in the EB should be estimated according to the characteristics in the survived neighboring blocks. In this work, the content of each EB is estimated firstly and then it is classified into one of the three categories. Where the three categories are defined as follows:

- Uniform block: the gray level of EB may be constant or nearly so. I.e., there is no obvious edge in the block.
- Edge block: the block locates on the boundary of two or more parts with different gray level. Because the size of block is not large, there are few edges passing through the block and the direction of each edge, in general, is with no or little change.
- Texture block: both gray level and edge direction varies significantly in the block, so the edge magnitudes of many directions are very strong.

From the above definition, it is a fact that the edge features play an important role in describing characteristics of content in local image. Therefore, the content in an error block can be estimated by estimating edge features in it. To determine the edge features in the EB, edges in the surrounding available blocks whose directions imply that they pass through the EB are computed [5]. Local edge gradient magnitude $g(i,j)$ and angular direction $\psi(i,j)$ at pixel (i,j) are calculated by convolving pixel values with Sobel Operator. The value of

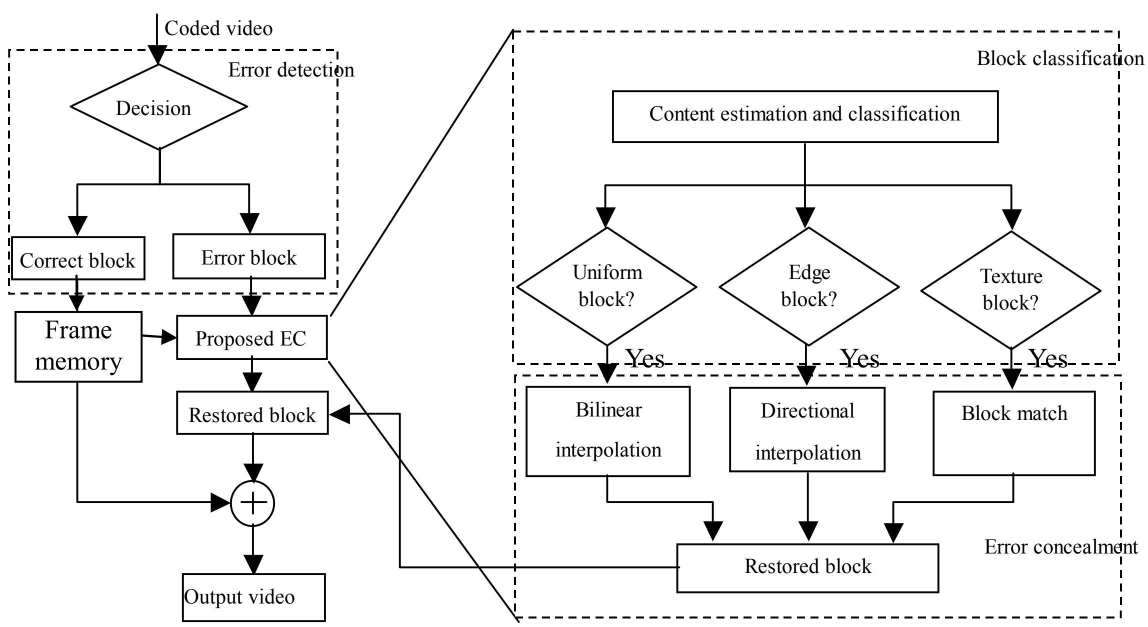


Fig. 2. Framework of the proposed EC algorithm

$\psi(i, j)$ is rounded to one of the eight directions equally spaced in the range from 0º to 180º. There is a counter d_k ($k = 0,1,\cdots,7$) for each of the eight directions. If the extension of the edge at pixel (*i,j*) belonging to neighboring blocks passes through the EB, the counter for that particular direction is incremented by the amount of $g(i, j)$. After all pixels of available neighborhood have been detected, the counters, d_ks, can be computed. And the likelihood of edges in each of the eight directions can be determined by the values of the counters.

Let $d_{\max} = Max\{d_1, d_2, \cdots, d_8\}$. To d_k, the edge of corresponding direction is defined as strong edge if $d_k > T \times d_{\max}$ and $d_{\max} \geq T_d$, where T and T_d are two certain thresholds. By trial and error, it is found that the value of T=0.55 and T_d=3000 work well for a large class of natural image. Denoting the number of directions with strong edges as N_d. According to the counters, the EB can be classified into three cases as follows:

Case 1—$d_{\max} < T_d$: In this case, the value of $d_{\max}$ is less than the threshold T_d. It is indicated that edges of all eight directions in the EB are not strong, i.e. the gray level of EB may be constant or nearly so. So the EB is a smooth block and can be considered as a *Uniform block*.

Case 2—$d_{\max} \geq T_d$ and $N_d \leq T_N$: This case means that there are strong edges within the block but the number of them is not more than a certain threshold, T_N. So the EB is considered as an *Edge block*.

Case 3—$d_{\max} \geq T_d$ and $N_d > T_N$: This case indicates that many directional or all directional edges in the EB are strong. Corresponding to Case 2, the block of this case can be considered as a *Texture block*.

IV. ERROR CONCEALMENT

In order to improve concealed image quality and speed up the concealment, each EB is reconstructed by a suitable method according to the categories it belongs to. In this work, they are concealed by bilinear interpolation (BI), directional interpolation (DI) and NMEC, respectively. And they are introduced in details in the following subsections.

A. Concealing Uniform blocks

To this kind of EB, the gray levels of pixels in the EB change slowly with the position. So each pixel in the EB can be concealed by linear interpolation using the nearest pixels from the four neighboring blocks along the block boundaries [2]. To an EB of size $N \times N$, pixel p(i,j), i,j=0,1,…,N-1, and it four nearest pixels, p(-1,j), p(N,j), p(i,-1) and p(i,N), are shown in Fig.3. The interpolated value of p(i,j) is interpolated by formula:

$$p(i, j) = \frac{p(-1, j)\times i + p(N, j)\times(N - i) + p(i,-1)\times j + p(i, N)\times(N - j)}{2N} \quad (1)$$

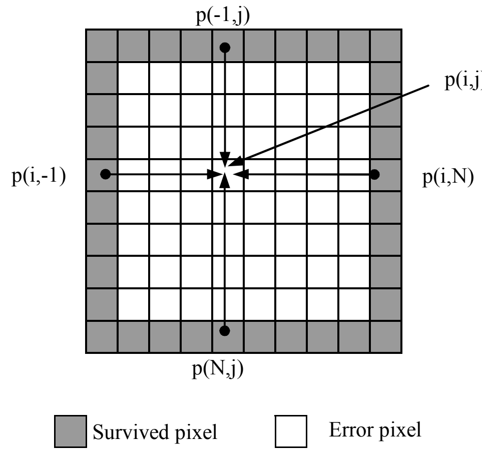


Fig.3. Conceal the Uniform block by bilinear interpolation

B. Concealing Edge blocks

Edges play an important role in the subjective image quality because the human visual system is sensitive to the structural information revealed by edges. Over smoothed edges blur an image while broken and falsely reconstructed edges cause unpleasant artifacts. Consistent edge profiles, i.e. consistent cross-edge sharpness and along-edge smoothness ensure consistent image structures as well as good visual quality. In order to preserve structural consistency, the EB classified as Edge blocks is interpolated along the edge direction. Therefore, if the edge of a certain direction is estimated as a strong edge, then the direction is selected as one interpolation direction and a series of one-dimensional interpolation are carried out along the direction to obtain pixel values within the EB. For example, the interpolation of pixel p (see Fig.4) along the given direction θ is

$$f_\theta(p) = \left[\sum_{x_n \in M} \frac{f_\theta(x_n)}{D_n^2}\right] \Bigg/ \sum_{x_n \in M} \frac{1}{D_n^2} \quad (2)$$

where M is the set of available neighboring pixels on the scanning line (passing through p and along direction θ). x_n is the nth pixel of M. D_n represents the distance from x_n to p.

If the number of strong edges is more than one, i.e. N_d >1, then the pixel values of EB are calculated from a weighted average of all directional interpolations. Denote S, a subset of $\{0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ, 157.5^\circ\}$, as the set of strong

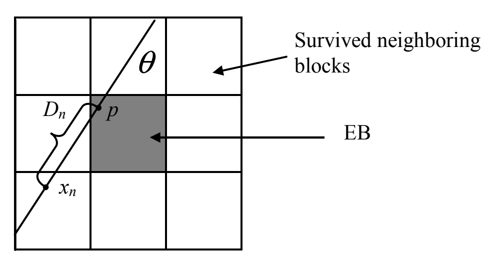


Fig.4. Conceal the Edge block by directional interpolation

edge directions. The value of pixel p, $f(p)$, is calculated as:

$$f(p)=\frac{\sum_{\theta_i \in S} d_i \times f_{\theta_i}(p)}{\sum_{\theta_i \in S} d_{\theta_i}} \tag{3}$$

Where θ_i is the directions $i \times 22.5^\circ$, $f_{\theta_i}(p)$ is the interpolated value along direction θ_i, and d_i is the same as that in section III.

C. Concealing Texture block

The spatial interpolation methods can obtain a well reconstructed image only when the damaged block is smoothly connected with its adjacent neighborhoods and the direction of edges passing through the EB should be with no change. However, because the damaged block with high details do not satisfy the predefined constraints which the interpolation methods rely on, the block concealed by interpolation will be blurred or cause much false stripes. In order to deal with the problem of reconstructing damaged blocks with high details, spatial similarity principles have been applied in existing neighboring blocks and concealment has been performed by searching the best match block [6][7][8]. The best match is the one that minimizes a cost function. The cost function is a function of the difference between the values of the one pixel-wide layer around the EB and the values of the one pixel-wide layer corresponding to one candidate block in the image.

Consider an $N \times N$ EB and a one pixel-wide layer around it (shown in Fig.5) in a damaged image, the cost function is written as:

$$V(s,t)=\frac{1}{K}\sum_{i=-1}^{N}\sum_{j=-1}^{N} w_{i,j} \times [f(s+i,t+j)-f(m+i,n+j)]^2 \tag{4}$$

where (m,n) and (s,t) are the up-left corner of the EB and the candidate block, respectively. $w_{i,j}$ is weighted coefficient, K is the number of compared pixel pairs. $w_{i,j}$ equals to 1 and K is incremented by 1 if both $(s+i,t+j)$ and $(m+i,n+j)$ are available or concealed pixels, and $w_{i,j}$ equals to 0 and K is incremented by 0 otherwise. In order to reduce the computational complexity and processing time, the searching range, i.e., given an EB, the part of the image is searched for best match blocks, is a $5N \times 5N$ square region with the EB at its center.

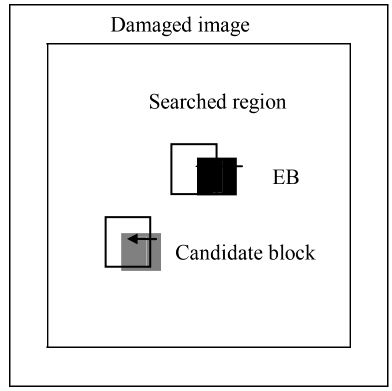


Fig. 5. Conceal the Texture block by NMEC

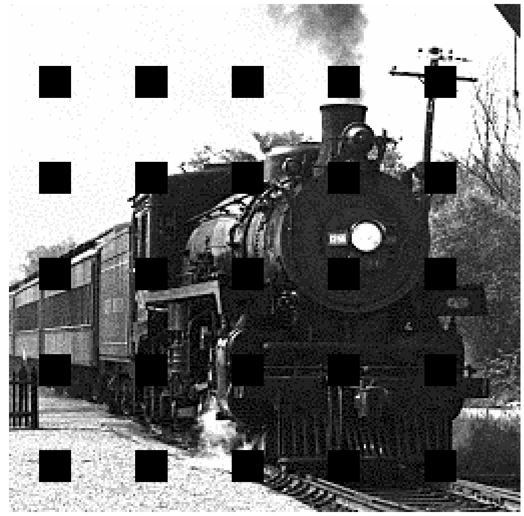
V. EXPERIMENTAL RESULTS

The 256×256 'Loco' image and test sequences 'Flower' with size 352×288 are used to evaluate the performance of the proposed algorithm. In the experiment, N=16, T=0.55, T_d=3000 and T_N =3. For comparison purposes, the result achieved by the proposed technique is matched with the results concealed by BI only, DI only and NMEC only, respectively.

The first test uses the four methods to perform EC under the condition that all damaged MBs are isolated. Fig.6 shows the experimental results. Fig.6(c) shows the image concealed only by BI. Horizontal or vertical edges, such as the EB on telegraph pole, can be recovered very well. However, other directional edges cannot be done and the method yields many horizontal and/or vertical strips passing through the concealed block, which are obvious in other concealed blocks. Fig.6(d) shows the restored image after the application of the DI. It can be seen that most of edges passing through the EB with different directions can be reconstructed while those blocks with high details are blurred. For example, the two concealed blocks locate in bottom-left. The result concealed by NMEC is shown in Fig.6(e), although the bottom-left two EBs are well masked, the horizon line within the EB locating on telegraph pole cannot be connected. By comparing, it can be seen obviously that the visual quality of image, which is concealed by the proposed algorithm, shown in Fig.6(f), is better than all images concealed by only one the three candidate methods.

The second experiment deals with the case that some damaged MBs are contiguous. Coded video information is grouped into packets, where each packet consists of coded data of a slice (22 blocks/slice, block size is 16×16). If a bit error happens, then the whole slice is considered as error, i.e. 22 consecutive blocks are EBs. The concealed results of an example damaged frame are shown in Fig.7. They are: (a) the original frame, (b) damaged image, (c) concealed only by BI, (d) concealed only by DI, (e) concealed only by BMEC and (f) reconstructed by the proposed algorithm. By comparing the concealed figures, the superior performance of the proposed algorithm becomes obvious. The results demonstrate that the proposed algorithm can obtain good subjective quality whether texture features of the EB is complicated or simple.

(a) Original frame



(b) Damaged image

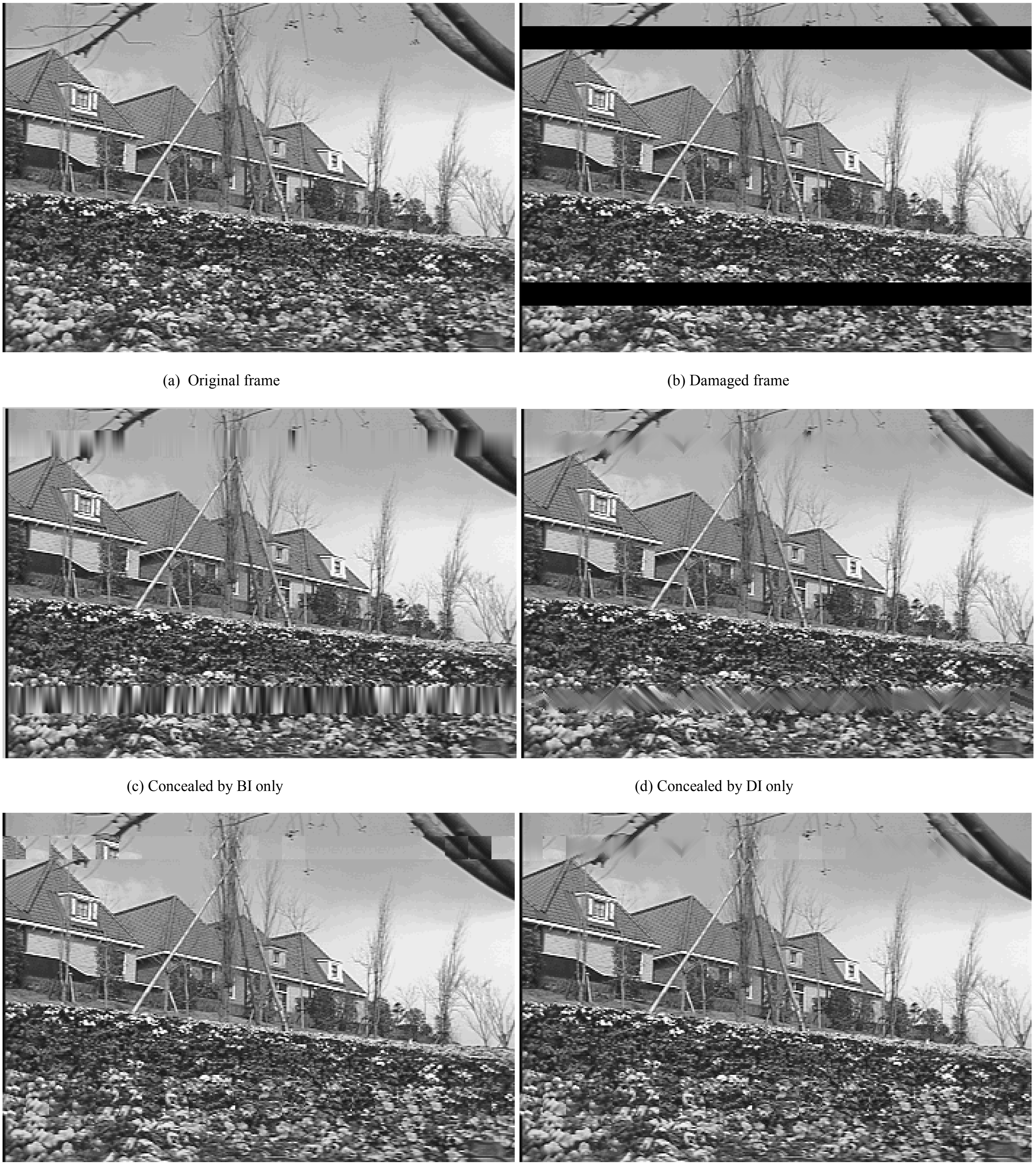
(c) Concealed by BI only

(d) Concealed by DI only

(e) Concealed by BMEC only

(f) Concealed by the proposed algorithm

Fig.6. A frame from the video sequence Flower



(a) Original frame

(b) Damaged frame

(c) Concealed by BI only

(d) Concealed by DI only

(e) Concealed by BMEC only

(f) Concealed by the proposed algorithm

Fig.7. A frame from the video sequence Flower

VI. CONCLUSION

In this paper, a content adaptive error concealment algorithm is proposed that reconstructs error blocks by taking advantages of both spatial interpolation EC and block matching EC. In order to select an appropriate EC, an effective content classifier is first presented. And then, to each category, a suitable EC method is introduced. Experimental results show that EBs with different contents are well reconstructed and the subjective visual quality concealed by the proposed algorithm is great improvement over that of concealed only by any one of the three candidate methods. The proposed algorithm exploits only the intra-frame information and is applicable to errors in MPEG, H.26x and JPEG coded images or videos transmitted over error-prone networks.

REFERENCES

[1] S.Shirani, F.Kossentini and R.Ward: 'Error Concealment Methods, A Comparative Study', Proceedings of IEEE Conference on Electrical and Computer Engineering, May,1999,pp.835-840

[2] J.W.Suh and Y.S.Ho: 'Error Concealment based on Directional Interpolation', IEEE Trans. Consumer Electronics, 1997,43, (3), pp.295-302.

[3] E.Asbun, P.Salama and E.J.Delp: 'Real-time Error Concealment in Digital Video Streams Using Digital Signal Processors', IEEE Trans. Consumer Electronics, 2001,47,(4),pp.904-909.

[4] X.Li and M.Orchard: 'Novel Sequential Error-Concealment Techniques Using Orientation Adaptive Interpolation', IEEE Trans. on Circuits and Systems for Video Tech., 2002,12,(10), pp. 857- 864.

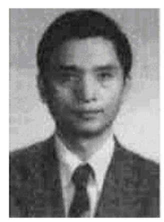
[5] W.Kwok and H.Sun: 'Multi-Directional Interpolation for Spatial Error Concealment', IEEE Trans. Consumer Electronics, 1993,39, (3), pp.455-460.

[6] W.Zeng and B.Liu: 'Geometric-Structure-Based Error Concealment with Novel Applications in Block-Based Low-Bit-Rate Coding', IEEE Trans. Circuits Syst. Video Technology, 1999, 9, (4), pp. 648-665.

[7] Z.Wang, Y.Yu and D.Zhang, 'Best neighborhood matching: an information loss restoration technique for block-based image coding systems', IEEE Trans. Image Processing, 1998,7,(7), pp.1056-1061.

[8] S.Tsekeridou and I.Pitas, 'MPEG-2 error concealment based on block-matching principles', IEEE Trans. Circuits Syst. Video Technology, 2000,10, (4), pp. 646-658

Zhang Rongfu was born in December 1971, in Anhui, China. He received the B.S degree in physics from Fuyang normal university, Anhui, China, in 1995, and the M.S. degree in optical information from University of Shanghai for Science and Technology, Shanghai, China, in 1998.
Since 1998, he was a Research Associate with the Department of electronics information, University of Shanghai for Science and Technology, Shanghai, China. He is currently working toward the Ph.D. degree in the Institute of Image Communication and Information Processing, Shanghai Jiaotong University, Shanghai, China. His current research interests are in the areas of video communication, HDTV, image processing and multimedia applications.



Zhou Yuanghua was born in 1940, in Jiangsu. He received the B.S degree in communication from Shanghai Jiaotong University, China, in 1965.
Since 1995, he is a professor at Shanghai Jiaotong University, China. In 1983-1984, he was also with New York state University as a visiting scholar, working on image processing. His research has involved image processing, video communication, HDTV and multimedia applications.

Huang Xiaodong was born in 1970, in Jiangxi, China. He received the B.S degree and the M.S. degree in physics from Nankai university, Tianjing, China, in 1994 and 1997, respectively. He is currently working toward the Ph.D. degree in the Institute of Image Communication and Information Processing, Shanghai Jiaotong University, Shanghai, China. His current research interests are in the areas of video communication, HDTV, image processing and multimedia applications.