ERROR RECOVERY FOR MPEG-4 SHAPE AND TEXTURE INFORMATION

Pei-Jun Lee and Liang-Gee Chen, Fellow, IEEE

Department of Electrical Eng.
National Taiwan University
Department of Information Management
St. John's & St. Mary's Institute of Technology
Phone: +886-2-23635251 ext 332, Fax:+886-2-23638247
Email:pjlee@video.ee.ntu.edu.tw

Graduate institute of Electronics Eng.
and Department of Electrical Eng.
National Taiwan University
Phone: +886-2-23635251 ext 233, Fax:+886-2-23638247
Email:lgchen@video.ee.ntu.edu.tw

ABSTRACT

In this paper, a new error concealment algorithm is proposed to repair damaged portions of the shape information and texture information. Based on Least Square Estimation method, the proposed algorithm can find the most similar block to replace the damaged block without worrying about the rotation, movement, or some luminance change between two sequential images. From the experimental results, the proposed algorithm achieves better performance than previous works especially on the case that the damaged portion includes the edge of object.

1. INTRODUCTION

It is well known that transmission of compressed visual information over unreliable networks cannot guarantee timely and lossless data delivery. In the current internet and wireless networks, unless the dedicated link can provide a guaranteed quality of service (QoS) which is available between the source and destination, data packets may be lost or corrupted due to either the traffic congestion or the noise from the physical channels. Error-free delivery of data packets can be achieved only by allowing retransmission of the lost or damaged packets through some mechanisms, for instance, automatic repeat request (ARQ). Channel encoding stage typically uses forward error correction (FEC) to protect them from transmission errors. At the receiver end, the packets are decoded by FEC and then unpacked. Finally the resulting bit stream is then input to the video decoder to reconstruct the original video. However, there may be some unacceptable delays in certain real-time applications. MPEG-4 is the latest ISO visual coding standard and has been applied to the 3rd generation International Mobile Telecommunication 2000 standard by ITU-T. MPEG-4 standard [1],[2] offers error resilience tools that help localization, isolation of the erroneous data and partial recovery of the remaining data [3], and it supports an object-based representation by allowing the coding of arbitrarily shaped video objects along with the texture and motion information. Therefore, the correctness of shape information is much more important than the other information. Video coding uses predictive coding standards and variable-length code (VLC) technique to reach lossless compression. However, Variable-length characteristic will make VLC sensitive to noise errors. Even a single error in a video bit-stream may have a large influence on the video quality because of the error propagation. Although previous works have utilized the synchronization codeword or packet to avoid error propagation, it is still worth improving their results to be much better. It is seen that without any error concealment there will be a serious error propagation occurring especially in the objects of the image. Therefore, the bit stream concealment for compressed transmission errors is very important to devise video encoding/decoding schemes. In video decoder, error concealment methods are used to recover the missing blocks to be acceptable subjectively. So far, lots of error concealment methods have been proposed [3][4][5]. However, those methods cannot find the most similar block when the image has some rotation, movement, or some luminance change between two sequential images. The error concealment in the texture information of the MPEG-4 video coded is similar to what has been vastly investigated in the standards of H.263 [8] and MPEG-2 [4][7][9]. [6] uses a defuzzification method to repair damaged portions of the shape information on arbitrarily shaped video objects. The weakness is that it cannot apply to conceal texture information. In this paper, we propose a new error concealment technique on the video decoder that can repair damaged portions of both shape information and texture information with the aids of Least-Square estimation method.

2. LEAST-SQUARE ESTIMATION FOR ERROR CONCEALMENT

In the data partition of MPEG-4 structure, shape data and motion vectors are most important. When the error occurs, all data in the bit stream between the point of error

0-7803-7690-0/02/$17.00 ©2002 IEEE

and the next resynchronization marker may be lost. Generally, when synchronization is lost, all the motion, shape, and texture data between two resynchronization words are discarded, thus, the decoder does not know the exact location of the error. MPEG-4 supports the data partition mode that allows the separation of all information in a video packet. The shape and motion data are represented by fewer bits than the texture data are represented. Therefore the impact of any errors in shape data is much more serious than in texture data.

In this section, we will propose a new error concealment technique in which Least Square Estimation (LSE)[10] plays an important role. Suppose n parameters x and measurements z are related as follows

$$z = Hx + v \tag{1}$$

The $(M \times n)$ matrix H is assumed known and the number of measurement (M) is at least as large as the number of unknown parameters (n), *i.e* $(M \geq n)$. In addition, H is with maximal rank n (except when noted otherwise). The vector v represents the uncertainty occurring in the measurement of x. The vector x denotes the unknown parameters to be estimated. z is assumed to have specific numerical values and these values are used in the estimator to obtain specific value, or estimates, of the parameters x. In LSE, the estimator is chosen to minimize the sum of the errors.

$$\varepsilon \equiv \frac{1}{2}\sum_{i=2}^{M} v_i^2 = \frac{1}{2}v^T v = \frac{1}{2}(z - Hx)^T (z - Hx) \tag{2}$$

we obtain the estimation of x

$$\hat{x}_{LS} = (H^T H)^{-1} H^T z \tag{3}$$

Based on LSE, we try to uses the side information of the damaged block to do the side matching in the transform domain. It is assumed that the location of the damaged block in an image can be detected. The damage may be in shape information or in the texture information. If only the shape information is damaged, then LSE can be used to recover the damaged block. Even both shape and texture information are damaged, LSE also can be applied to both of them respectively. Finally, two estimated blocks by LSE are combined to recover the damaged block. Let us use Fig. 1 to explain the proposed technique. The damaged block $X_{i,j}$ in an image is marked with "A". Those pixels, which are adjacent to and surround the damaged block, are marked with "B" (it is the side information of the damaged block). Let the pixels in B be denoted as follows

$$X_{i,j\_s} = [X_u{}^T, X_b{}^T, X_l{}^T, X_r{}^T]^T$$

where X_u, X_b, X_l, and X_r are the upper, under, left, and right side information of the damaged block, respectively. Similarly, let $X_{k,l\_s} = [X_{uc}{}^T, X_{bc}{}^T, X_{lc}{}^T, X_{rc}{}^T]^T$ denote the candidate in the previous frame, where X_{uc}, X_{bc}, X_{lc}, and X_{rc} are the upper, under, left, and right side information of the candidate block $X_{k,l}$, respectively. X_{k,l_s} will be caught in a limited region "C" of the temporal domain (see Fig. 2). We expect there is a relationship between X_{i,j_s} and X_{k,l_s} as follows

$$X_{i,j\_s} = H_{k,l\_s} \cdot T_{k,l} \tag{4}$$

Any one element of X_{i,j_s} according to (4) can be represented by

$$x_{i,j\_s} = a_0 + a_1 x_{k,l\_s} + a_2 x_{k,l\_s}^2 + \ldots + a_n x_{k,l\_s}^n \tag{5}$$

That is, $T_{k,l}$ denotes the unknown parameters as

$$T_{k,l} = [a_0, a_1, \ldots a_n]^T \tag{6}$$

Now, by LSE, $T_{k,l}$ can be obtained by (3)

$$T_{k,l} = (H_{k,l\_s}^T \cdot H_{k,l\_s})^{-1} \cdot H_{k,l\_s}^T \cdot X_{i,j-s} \tag{7}$$

Then the obtained $T_{k,l}$ can be used to transform $X_{k,l}$ to $\hat{X}_{i,j}$ as

$$\hat{X}_{i,j} = H_{k,l} \cdot T_{k,l}, \tag{8}$$

It should be noted that in (7), X_{i,j_s} is fixed, but H_{k,l_s} and $T_{k,l}$ depend on the caught location in the region C. Since there may be some rotation, movement, or luminance change between two sequential images, in order to find the most similar block $\hat{X}_{i,j}$, we should choose the best $T_{k,l}$ corresponding to different $X_{k,l}$ such that the error $\varepsilon = \sum_{i,j} \left| \hat{x}_{i,j\_s} - x_{i,j\_s} \right|$ is minimum. Finally, the optimal $T_{k,l}$ is obtained, by (8) the most similar block $\hat{X}_{i,j}$ is obtained also. Consequently, $\hat{X}_{i,j}$ can be utilized to replace the damaged block $X_{i,j}$.

When the error is detected in the shape information, then LSE can be used to conceal the damage block. If the shape and motion information are correct, the error occurs in texture information only, then we can use motion information to conceal texture information. Unfortunately, if the motion information is lost also, LSE still can be applied to conceal the damage texture information. Let us summarize the above operations to the following produce.

Step 1: Detect the damaged block "A" and determine the surrounding and adjacent pixels "B". Moreover, determine the research region "C" in the temporal domain.

Step 2: Establish several $T_{k,l}$ corresponding to the different location (k,l) in the region C.

Step 3: Find the $\hat{x}_{i,j-s}$, which has the smallest error ε, and then obtain the corresponding matrix $H_{k,l}$ and the vector $T_{k,l}$.

Step 4: Transform the block information by (8) to get a new block $\hat{X}_{i,j}$ to replace the damaged block $X_{i,j}$.

The above procedure can be applied to either shape or texture information. Moreover, it can be applied to the spatial domain also just the research region to be changed. For reducing computation complexity, we can simplify step 2 and step 3 by choosing $X_{k,l}$ in region C at the same corresponding location as $X_{i,j}$ in block A. Thus, only one $T_{k,l}$ is needed in step 4.

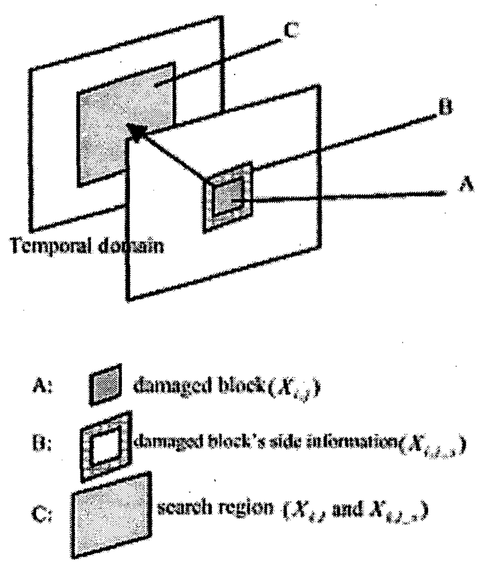

Fig.1 Definition of the search range and side information

3. EXPERIMENTAL RESULTS

This section gives some experiments of error concealment for various types of video sequences. There are five methods to be evaluated_in those experiments. **1. No concealment:** Not any concealment methods applied in this case. **2. Frame replacements or zero replacement (ZR)**: Each lost pixel is replaced by the pixel at the corresponding location in the previous frame. **3. Above motion vector (AMV)**: The motion vector from the marcoblock above the lost macroblock is used to perform motion-compensated concealment. **4. Direct matching method (DMM):** The block in the same corresponding location in the temporal domain (in the previous frame) with the damaged block is selected to replace the damaged block directly. That is the special case of LSE with $a_1 = 1, a_i = 0, i = 0, 2, 3..., n$. This case is as the same as using the motion vector to cancel the lost block. **5. First order Least-square estimation method (LSE):** That is the proposed method. In the temporal domain, the size of the damaged block is 16×16 (see Fig. 1) "A". Side information is 1 pixel of surrounding damaged block (see Fig. 1) "B" and the search region is 20×20 in the previous frame (see Fig. 1) "C".

The video sequences in the experiments are "Children" and "Frank". "Children" sequences have 150 frames and each frame format is 176×144 (QCIF) in luminance. "Frank" sequence has 80 frames and each frame format is 320×240.

Tables 1~2 show the objective view of the performance comparison among the five methods described above for "Children" and "Frank" sequences with 5%, 6%,..., 10% packet lost blocks for the shape information and texture information, respectively. PSNR, PSE and PER denote the peak-to-peak signal and noise ratio, the pixel of shape error, and the pixel error rate, respectively. Note that PER can be obtained by PER= (PSE / total pixel of damage)*100%.

Fig.2 shows the 26^{th} frame from the "Children" sequence combined the shaped and texture information, in which a large number of errors can be seen along the ball. Fig.3 shows the 68^{th} frame from the "Frank" sequence combined the shaped and texture information, in which a large number of errors can be seen along the head. In Figs. 2 and 3, (a) is the original picture. (b) is the damaged picture with 10% lost blocks. (c), (d), (e), and (f) are the recovered pictures by ZR, AMV, DMM and LSE, respectively.

3. CONCLUSIONS

This paper has presented a new error concealment method to repair the damaged portion of the shaped information and texture information. This method is based on Least-Square Estimation method with side matching. The experimental results indicate that there is a significant improvement comparing with those existing algorithms in terms of both subjective and objective evaluation. Several video sequences have been tested with three different methods. It has been shown that the proposed method has the best performance and will be suitable to provide robust video coding for video communication with/without objects.

REFERENCES

[1] MPEG-4 overview, http://drogo.cscelt.stet.it/mpeg/standards/MPEG-4.htm, July 1998.

[2] B.Erol, A. Dumitras, and F. Kossentini, " Emerging MPEG Standards: MPEG-4 and MPEG-7," in *Handbook of Image and Video Processing*: Academic Press, 2000.

[3] Y.Wang, Q.F.Zhu, And L.Shaw, " Maximally smooth image recovery in transform coding," *IEEE TRANS, COMMUN.*, vol.41, pp.1544-1551, Oct.1993.

[4] *Coding of audio-visual object: Video,* ISO/IEC JTC!/SC29/WG11, Jan. 1999

[5] Zhou Wang, Yinglin Yu, and David Zhang, " Best neighborhood matching: an information loss restoration technique for block-based image coding systems" *IEEE Tran. Image Processing,* Vol. 7, No. 7, July 1998.

[6] P.J Lee, L.G Chen, W.J Wang, and M.J Chen, " Robust error concealment algorithm for MPEG-4 with the aids of fuzzy theory," ICCE, June 2001.

[7] Y. Wang and Q. Zhu, " Signal loss recovery in DCT-based image and video codecs," in *SPIE Proc. Vis. Commun. Image Processing*, vol. 1605, pp. 667–678, Nov. 1991.

[8] H. Sun and W. Kwok, " Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Trans. Image Processing*, vol. 4, pp. 470–477, Apr. 1995.

[9] S. S. Hemami and T. H. Y. Meng, " Transform coded image reconstruction exploiting interblock correlation," *IEEE Trans. Image Processing*, vol. 4, pp. 1023–1027, July 1995.

[10] H.W. Sorenson, Parameter Estimation Principles and Problems, Control and Systems Theory, Vol. 9, 1985.

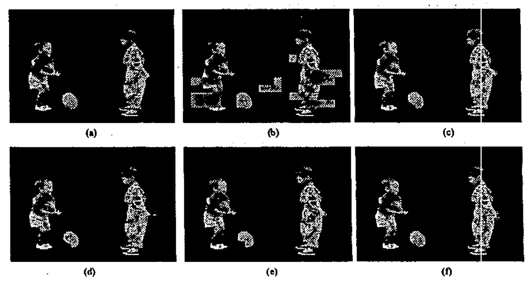

Fig.2 Reconstructed pictures for Children Sequence combining the shape and texture information. (a) The original picture, (b) The damaged picture, (c) The recovered picture by zero replacement, (d) The recovered picture by Above Motion Vector, (e) The recovered picture by Direct matching method, (f) The recovered picture by the proposed method (LSE).

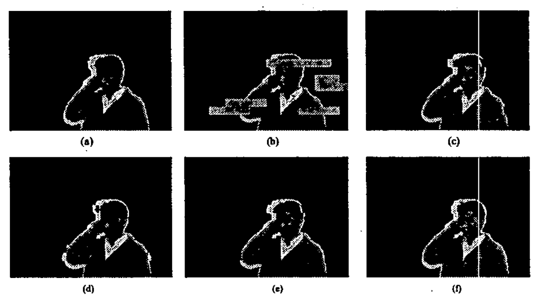

Fig.3 Reconstructed pictures for Frank Sequence combining the shape and texture information. (a) The original picture, (b) The damaged picture, (c) The recovered picture by zero replacement, (d) The recovered picture by Above Motion Vector, (e) The recovered picture by Direct matching method, (f) The recovered picture by the proposed method (LSE).

Table.1 Comparison of the texture information with packet loss

Sequence	Children					Frank				
Methods	NC	ZR	AMV	DMM	LSE	NC	ZR	AMV	DMM	LSE
Error rate	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR
0.05	17.36	28.26	28.91	30.08	30.87	11.32	26.69	29.27	30.08	33.76
0.06	16.44	24.49	26.86	27.92	28.73	11.32	26.65	29.20	29.91	32.37
0.07	16.37	23.15	26.21	26.31	27.18	11.31	26.50	28.97	29.72	32.21
0.08	16.14	22.76	25.82	26.05	26.83	11.31	25.90	28.57	29.23	31.42
0.09	16.05	22.62	25.45	25.92	26.79	11.30	24.32	26.98	29.18	31.12
0.10	15.89	22.31	25.38	25.76	26.58	11.29	23.76	26.22	29.12	30.70
Average	16.38	24.67	26.44	27.01	27.83	11.31	25.64	28.20	29.54	31.93

Table.2 Comparison of the shape information of Children sequence (QCIF) with packet loss

Methods	NC			ZR			AMV			DMM			LSE		
error rate	PSNR	PSE	PER%	PSNR	PSE	PER%	PSNR	PSE	PER%	PSNR	PSE	PER%	PSNR	PSE	PER%
0.05	6.04	1280.00	100.00	18.78	34.33	2.68	18.19	35.76	2.79	18.55	33.29	2.60	20.91	20.95	1.64
0.06	6.04	1536.00	100.00	18.25	35.51	2.31	17.87	38.75	2.52	18.13	34.79	2.26	20.63	24.05	1.57
0.07	6.04	1792.00	100.00	17.59	35.71	1.99	17.25	39.77	2.22	17.50	35.46	1.98	20.04	25.53	1.42
0.08	6.03	2048.00	100.00	17.12	36.04	1.76	17.00	42.68	2.08	17.26	36.85	1.80	19.34	27.74	1.35
0.09	6.03	2304.00	100.00	17.06	54.36	2.36	16.81	57.15	2.48	17.20	54.26	2.36	18.39	46.51	2.02
0.10	6.03	2560.00	100.00	16.84	58.72	2.29	16.66	63.23	2.47	16.75	58.34	2.28	18.10	51.46	2.01
Average	6.03	1920.00	100.00	17.61	42.44	2.23	17.29	46.22	2.43	17.56	42.16	2.21	19.57	32.71	1.67