Error detection using mode information of macroblocks for video transmission

Yong Fang (方 勇), Chengke Wu (吴成柯), and Hong Zhu (朱 红)

National Key Laboratory on ISN, Xidian University, Xi'an 710071

Received March 28, 2005

A new method for error detection using mode information of macroblocks (MBs) is proposed. For decodable inter MBs, inter residues are calculated by adding up absolute values of received residual pixels and intra complexities are estimated by that of motion compensated reference blocks. If inter residues are larger than intra complexities by a predefined quantity, MBs are considered to be erroneous. For decodable intra MBs, the connective smoothness of the current MB with correctly decoded neighboring MBs is tested to find erroneous MBs. Combined with error concealment, the new method improves the quality of reconstructed images by about 0.5—1 dB in peak signal-noise ratio (PSNR).

OCIS codes: 100.0100, 280.0280.

When video signals are transmitted over noisy channels such as wireless networks, there will be errors in the transmitted bit stream. Traditional coding methods make video more sensitive to channel errors because of spatial and temporal propagation due to using variable length codes (VLC) and motion compensation, respectively. Error detection and concealment are extremely important components of error robust video codec. The process of error detection can be divided into two steps. First, video bit stream is parsed to find whether there is any error. If any, further behaviors are carried on to determine the positions of errors. Many techniques have been developed for the latter purpose[1−3]. However, these methods are complex and not suited to the present hybrid motion estimation (ME), discrete cosine transform (DCT) and VLC framework.

In this letter, a new method for error detection is proposed. The new method aims to remove erroneously decoded macroblocks (MBs) based on the prerequisite that header and motion information are received correctly. As for how to detect and conceal errors in header and motion information, please refer to Ref. [3].

Firstly, a frame is divided into MBs. For each MB $(\mathrm{orig}(i,j))$, ME is performed to find the optimal reference block $(\mathrm{ref_{opt}}(i,j))$. The minimized inter residue $\mathrm{SAD_{orig}}$ is calculated as

$$\mathrm{SAD_{orig}} = \sum_{i=0,j=0}^{15,15} \mathrm{abs}(\mathrm{orig}(i,j) - \mathrm{ref_{opt}}(i,j)).$$

Then, the intra complexity (MB_Comp) is calculated as

$$\mathrm{MB\_Comp} = \sum_{i=0,j=0}^{15,15} \mathrm{abs}(\mathrm{orig}(i,j) - \mathrm{MB\_DC}),$$

where MB_DC is the direct current (DC) component of the current MB.

Intra mode is chosen if $\mathrm{MB\_Comp} < (\mathrm{SAD_{orig}} - C)$ (where C is a const and set to 512 in Ref. [5]), otherwise inter mode is chosen. Depending on MB modes, either original pixels or residual pixels are encoded. We add up the absolute values of the local reconstructed residual pixels $(\mathrm{diff_{enc}}(i,j))$ and denote the sum as $\mathrm{SAD_{enc}}$. Because the coding process is lossy, we have

$$\mathrm{SAD_{enc}} < \mathrm{SAD_{orig}} < \mathrm{MB\_Comp} + C. \quad (1)$$

In Fig. 1, $\mathrm{SAD_{enc}}$ and $\mathrm{SAD_{orig}}$ of MBs of the second frame of "Foreman" are plotted.

After one inter MB is decoded, inverse discrete cosine transform (IDCT) is executed to obtain residual pixels $\mathrm{diff_{dec}}(i,j)$. $\mathrm{diff_{dec}}(i,j)$ can be written as

$$\mathrm{diff_{dec}}(i,j) = \mathrm{diff_{enc}}(i,j) + \mathrm{err}(i,j),$$

where $\mathrm{err}(i,j)$ is the error image introduced by channel noise. We add up the absolute values of $\mathrm{diff_{dec}}(i,j)$ and denote the sum as $\mathrm{SAD_{dec}}$. If the channel is error-free, we have

$$\mathrm{SAD_{dec}} = \mathrm{SAD_{enc}} < \mathrm{SAD_{orig}} < \mathrm{MB\_Comp} + C,$$

i.e.,

$$\mathrm{SAD_{dec}} < \mathrm{MB\_Comp} + C. \quad (2)$$

Otherwise, $\mathrm{SAD_{dec}}$ will take a different values from $\mathrm{SAD_{enc}}$. The difference between $\mathrm{SAD_{dec}}$ and $\mathrm{SAD_{enc}}$ is determined by $\mathrm{err}(i,j)$. There exist two cases as follows. (a) If high frequency components of the current MB are

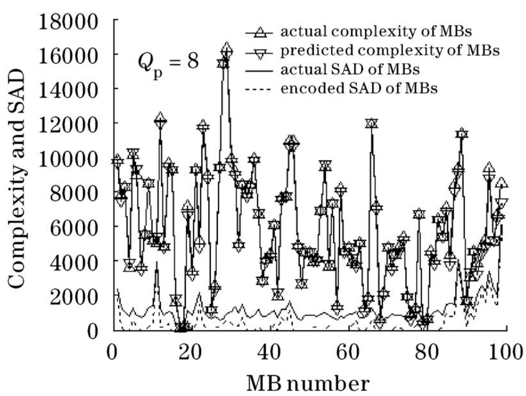

Fig. 1. Intra complexities and inter residues.

lost, $SAD_{dec} < SAD_{enc}$; (b) if extra high frequency components are decoded erroneously, $SAD_{dec} > SAD_{enc}$.

These two cases have different impacts on reconstructed images. For case (a), the impact is often trivial and can be tolerable. However for case (b), the reconstructed image may be degraded seriously. For this reason, our interest will be focused on the case (b) hereinafter.

For case (b), serious errors cause large SAD_{dec} usually. The Eq. (2) will not hold if SAD_{dec} is too large. Intuitively, we hope that Eq. (2) can be used as the criterion whether a MB has been decoded correctly. However because the coding process is lossy, the exact original MB cannot be reconstructed at decoders. To get around this problem, we use the motion compensated reference block $MC\_Ref(i, j)$ to estimate MB_Comp. We denote the intra complexity of the $MC\_Ref(i, j)$ as MB_Comp_{ref}. Simulation results show that MB_DC_{ref} is a good approximate to MB_Comp (see Fig. 1).

On substituting MB_DC_{ref} for MB_Comp, we have

$$SAD_{dec} < (MB\_Comp_{ref} + C'), \qquad (3)$$

where C' is a const, and experiments confirm that when $C' = C$, good results will be obtained.

Equation (3) is used as the criterion whether an inter MB has been decoded correctly. For any decodable inter MB, if Eq. (3) holds, it is considered to be decoded correctly. Otherwise, it is deemed to be erroneous. As for the case that MBs are decoded erroneously while Eq. (3) still holds, the errors should not be so serious and can be tolerable.

It deserves to be pointed out that two traditional methods dealing with damaged video packets can also be seemed to be special cases of Eq. (3). When we let C' be a maximum, Eq. (3) will hold always, which corresponds to the case that all decodable MBs are deemed to be correct. On the contrary, if we let C' be a minimum, Eq. (3) will never hold, which corresponds to the case that all MBs are discarded.

Because there are many fewer intra MBs in video bit stream, error detection of intra MBs is not so important an issue as that of inter MBs. In our simulation, DCs of decodable MBs are compared with that of spatially neighboring MBs and edge pixels are used to test the connective smoothness of the current MB with neighboring MBs. If there is a large difference between the current MB and neighboring MBs, the current MB is deemed to be erroneously decoded.

In our simulation, three different schemes are used to deal with damaged video packets: all MBs are concealed (no error detection, NED) (i.e, if any error is detected in one packet, all MBs in the packet are concealed), only undecodable MBs are concealed (no error detection and remain all decodable MBs, NER), and detecting damaged MBs using proposed method and only correctly decoded MBs are remained (error detection, ED). For intra MBs, spatial concealment algorithm is based on weighted pixel averaging. For inter MBs, motion compensated reference blocks are copied directly to conceal damaged MBs. Three standard quarter common intermediate format (QCIF) video sequences are tested: "Foreman", "Coastguard" and "Container", with the frame rate 25 frames/s. According to the recommendation of MPEG-4 verification model, C is set to 512 at the encoder when making decision on coding modes of MBs (for a common MPEG-4 encoder, this recommended value should be adopted). For every 50 frames, one I frame is inserted. Each video packet contains 11 MBs. Header information (include coding modes of MBs) and textural information of MBs are contained in different packets. Header information is protected to a higher degree. If any error is detected in header information, all MBs in the same packet are taken as inter MBs for P frames (All MBs must be intra for I frame, so this case is not considered furthermore).

In Tables 1, 2, and 3, the average luminance peak signal-noise ratio (PSNR) (dB) comparisons are listed for different bit error rates (BERs) when video bit stream is transmitted across channels with different bandwidths. Channel bandwidths in Tables 1, 2, and 3 are 64, 96, and 128 kb/s, respectively. We find that about 0.5—1 dB improvement can be found in our ED method compared with the NER method. Compared with the NED method, the average PSNR of the ED method is also improved.

Table 1. Objective Image Quality Comparison of Three Schemes (64 kb/s)

BER	Foreman			Coastguard			Carphone		
	NED	NER	ED	NED	NER	ED	NED	NER	ED
1.35×10^{-3}	22.21	22.12	22.65	20.33	19.50	20.55	25.41	25.00	25.72
1.12×10^{-4}	28.85	28.85	29.21	25.91	25.61	26.14	29.62	29.47	29.74
8.77×10^{-5}	29.11	28.92	29.23	26.40	26.04	26.53	29.73	29.92	29.90

Table 2. Objective Image Quality Comparison of Three Schemes (96 kb/s)

BER	Foreman			Coastguard			Carphone		
	NED	NER	ED	NED	NER	ED	NED	NER	ED
1.35×10^{-3}	24.10	24.03	24.60	22.26	21.49	22.68	27.45	27.13	27.54
1.12×10^{-4}	30.94	30.37	31.14	27.85	27.83	28.17	31.59	31.40	31.63
8.77×10^{-5}	31.15	30.41	31.31	28.31	28.13	28.65	31.63	32.01	31.86

Table 3. Objective Image Quality Comparison of Three Schemes (128 kb/s)

BER	Foreman			Coastguard			Carphone		
	NED	NER	ED	NED	NER	ED	NED	NER	ED
1.35×10^{-3}	26.23	26.01	26.71	24.34	23.59	24.63	29.51	29.05	29.68
1.12×10^{-4}	32.97	32.87	33.10	29.96	29.76	30.19	33.67	33.38	33.71
8.77×10^{-5}	33.13	33.00	33.21	30.44	30.11	30.67	33.80	34.01	33.89

Fig. 2. Subjective image quality comparison of the 33rd frame of "Foreman".

Figure 2 shows the 33rd decoded frame of "Foreman" using the NED, NER, and ED methods when BER is 1.35×10^{-3}, respectively. Similar conclusion to that of objective comparisons can be obtained. The reconstructed image quality of the ED method is better than that of other two methods. In these three methods, the image quality of NER method is the worst because of the existence of erroneously decoded MBs.

A novel error detection method using mode information of MBs is proposed. The method can remove those MBs with serious texture errors. Experiments show that combined with error concealment, the new method improves reconstructed image quality. In addition, our method is simple and easy to apply.

This work was supported by the National Natural Science Foundation of China (No. 60372043, 60272050). Y. Fang's e-mail address is yfang79@gmail.com or fangyong79@hotmail.com.

References

1. K. N. Ngan and R. Stelle, IEEE Trans. Commun. **30,** 257 (1982).
2. K. M. Rose and A. Heiman, IEEE Trans. Commun. **37,** 373 (1989).
3. J. Wu and C. Wu, Chin. Opt. Lett. **2,** 325 (2004).
4. W. M. Lam, A. R. Reibman, and B. Liu, IEEE Int. Conf. Acoustics, Speech, Signal Proc. **5,** 417 (1993).
5. Y. Wang and Y. Q. Zhang, in *Video Processing and Communications* (Tsinghua University Press, Beijing, 2003) chap.9, p.299.