AN ERROR-RESILIENT H.263+ CODING SCHEME FOR VIDEO TRANSMISSION OVER WIRELESS NETWORKS[1]

Li Jian[*] Bie Hongxia[* **]

[*](*School of Information Engineering*, *Beijing University of Posts & Telecommunications*, *Beijing 100876*, *China*)

[**](*Northeast Institute of Geography and Agricultural Ecology*, *CAS*, *Changchun 130012*, *China*)

Abstract Video transmission over wireless networks has received much attention recently for its restricted bandwidth and high bit-error rate. Based on H.263+, by reversing part stream sequences of each Group Of Block (GOB), an error resilient scheme is presented to improve video robustness without additional bandwidth burden. Error patterns are employed to simulate Wideband Code Division Multiple Access (WCDMA) channels to check out error resilience performances. Simulation results show that both subjective and objective qualities of the reconstructed images are improved remarkably. The mean Peak Signal to Noise Ratio (PSNR) is increased by 0.5dB, and the highest increment is 2dB.

Key words Wireless networks; Error detection; Error resilient; Error localization; Resynchronization

I. Introduction

Multimedia applications over wireless networks are becoming attractive, which are often bandwidth limited[1,2] and noisy[1,3,4]. Many video coding standards, such as H.263[5] and Moving Picture Expert Group (MPEG) family have been proposed. Many studies have been carried out for higher compression rate along with high efficiency by using Variable Length Coding (VLC). Traditional codec schemes are not robust in error prone circumstances for VLC errors could decrease the quality of video clips seriously[1,2].

Real-time video transmission, which requires high bandwidth, low probability of channel error and short transmission delay, is the most attractive technology for many applications, such as video-telephony and video surveillance systems. While the strict time limitation is quite difficult to meet in mobile networks. Wireless channels suffer from serious bit error and restrained bandwidth[1,2], which make video communication over wireless channels more challenging. For real-time video transmission in mobile system, the codec must be error resilient and efficient.

Transmission errors in a mobile channel range from single bit errors to burst errors or even intermittent loss of connection[3]. There are two basic categories of error correcting codes to combat transmission errors, Forward Error Correction (FEC) and closed-loop error control techniques like Automatic Repeat reQuest (ARQ) at the expense of sacrificing transmission bandwidth or adding up transmission delay. In FEC, redundancy is added to data packets before transmission and is used by the receivers to detect and correct errors[1]. The widely varying error conditions in mobile communication channels make FEC useless because FEC would lead to a prohibitive amount of redundancy in the worst cases, which introduces a high transmission overhead. ARQ has been shown to be more effective than FEC[3]. However, retransmission of corrupted data frames introduces additional delay, which might be unacceptable for real-time conversational services. Another problem of ARQ is that it needs a feedback channel[2].

The paper aims to improve H.263 codec scheme for reliable video transmission via mobile communication channels. H.263 is a compression standard to transmit video clips using phone networks at the data rates less than 64kbit/s[5]. H.263+ and H.263++ are the extension of H.263, aiming to improve the compression efficiency and to broaden the range of application. A typical H.263 video stream is composed of an intra-frame (I frame) and several inter-frames (P frame). An I frame is coded independently without any previous reference frame and can be decoded by itself. A P frame is composed of the difference between the current image and the last coded I frame or P frame[5]. The compressed result is extremely vulnerable against transmission errors, but low bit-rate video coding schemes rely on P frame coding for high coding efficiency. And they use the previous encoded and reconstructed video frames to predict the next frame. Therefore, the loss of information in one frame has considerable impact on the following frames. Since some residual transmission errors inevitably corrupt video bit stream, this vulnerability precludes the use of low bit-rate

---

[1] Manuscript received date: April 11, 2005;
revised date: September 19, 2005.
Communication author: Li Jian, born in 1978, male, Master candidate. School of Information Engineering, Mailbox 261, Beijing University of Posts and Telecommunications, Beijing 100876, China.
lighter_lj@163.com.

video coding schemes designed for error-free channels without special measures. So some measures have to be built into video encoding and decoding algorithms themselves and form the “last line of defense” if FEC and ARQ fail.

H.263 employs VLC to achieve high compression rate. No delimiters are used in such entropy coder to separate the successive binary code words. And such codes are highly susceptible to transmission errors. Even a single bit error may cause significant changes in the decoded symbol stream[1]. Once video-frame synchronization fails in decoding, video degradation will continue until synchronization is reestablished. Some protection measures can be obtained by wrapping VLC in an error correcting code[1,6], but the additional redundancy defeats the original purpose of bit reduction using entropy code. In this paper, an improved coding scheme based on H.263+ baseline is proposed to resist the errors from wireless channels, hardly sacrificing any transmission bandwidth.

II. Improved H.263+ Codec Scheme

In real-time wireless surveillance systems, Quarter Common Intermediate Format (QCIF) (176×144 pixels) resolution is considered. The QCIF image can be divided into 99(11×9) Macro Blocks (MBs), which is comprised of four 8×8 luminance blocks and two 8×8 chrominance blocks. A fixed number of successive MBs are usually grouped into a Group Of Blocks (GOB), for example, 11MBs for QCIF images. H.263 inserts some synchronization makers at the frame imperatively and at the GOB layers optionally. The latter is often used to resynchronize when errors occur. In this paper the GOB synchronizing markers are assumed to be inserted at the beginning of each MB row.

In the Annex V of H.263++, a variation of Huffman code called Reversible VLC (RVLC)[6] has been proposed. It was shown that RVLC could offer the same coding efficiency as all the other Huffman coded while imposing a symmetry structure of each VLC word essentially. And each RVLC code can be decoded in the forward and the backward directions. Therefore, if an error is detected during RVLC decoding, the decoder can jump to the end of the code stream and start decoding backwards. As a result, only part data actually hit by the errors is lost, and more data can be recovered in RVLC. Thus Annex V will greatly enhance the performance of H.263++ in an error-prone environment. But at the same time, RVLC changes the structure of H.263+ baseline, increases the complexity of codec for data partition and sacrifices the efficiency due to the overhead maker words. In addition, Discrete Cosine Transform (DCT) coefficients are run-length coded and can not be protected by RVLC[4,6]. And when multiple bits, which are further apart, are lost, the entire segments between them cannot be recovered. So the major disadvantages of RVLC include decreased compression efficiency, increased error propagation and increased decoding complexity.

In H.263+ baseline, the error diffusing is serious and a single bit transmission error would make the decoder lose synchronization words or even block it. So it needs to search the next synchronization words when error bits are detected. The data between the error bits and the next resynchronization words will be discarded unless some remedial measures are taken.

Referencing RVLC, H.263+ baseline can also be improved by back-direction decoding. As illustrated in Fig.1, in the improved mode the coded MBs bits of the GOB back part are reversed while encoding. These MBs can be decoded in backward direction from the next resynchronization words if an error occurs to stop the fore-direction decoding. In essence, for the bit-stream in an MB there are resynchronization words not only after it but also before it. All the coefficients of the back half MBs in the GOB can be decoded in backward direction and the coding efficiency does not decrease. Combined with error concealment, the improved codec scheme can greatly improve the quality of transmitted images over noisy channels, such as wireless networks.

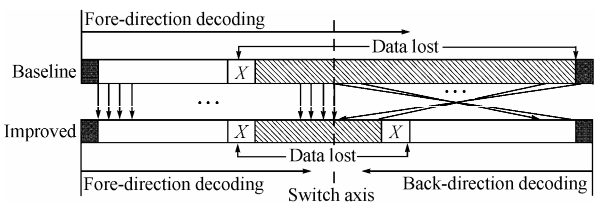

Fig.1 Comparison between baseline and the improved codec scheme

The improved method employs GOB header (the shaded bit in Fig.1) to prevent error diffusing in successive GOBs and makes full use of synchronization words to do double-direction decoding.

III. Improved Encoder and Decoder

The stream structures of H.263+ baseline and the improvement are compared in Fig.2. And the difference is that the bits of back part GOB in the improved codes are reversed bit by bit. No overhead codes are introduced in the improved algorithm, and it is robust against errors without sacrificing efficiency and complexity. The detailed process of encoding is illustrated in Fig.3.

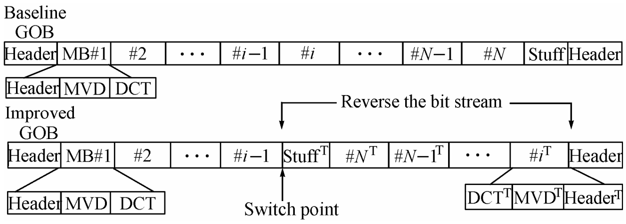

Fig.2 Comparison between baseline and the improved GOB structures

Note: MVD is the difference between the actual MV and the predicated MV. And symbol T means reversion of bit stream, namely bit stream is re-written from right to left.

Fig.3 shows that the encoding process is not different from baseline for the first half MBs of the current GOB(GOB p1 in Fig.3). The bit position of the end bit of GOB p1, called switch point, is denoted as $i-1$. The encoder continues working until the whole GOB has been encoded according to baseline and records the end bit position of the GOB as k. Some padding bits may be included before the position k for memory alignment. Then the bits from $i-1$ to k are reversed bit by bit and re-written back to the same buffer.

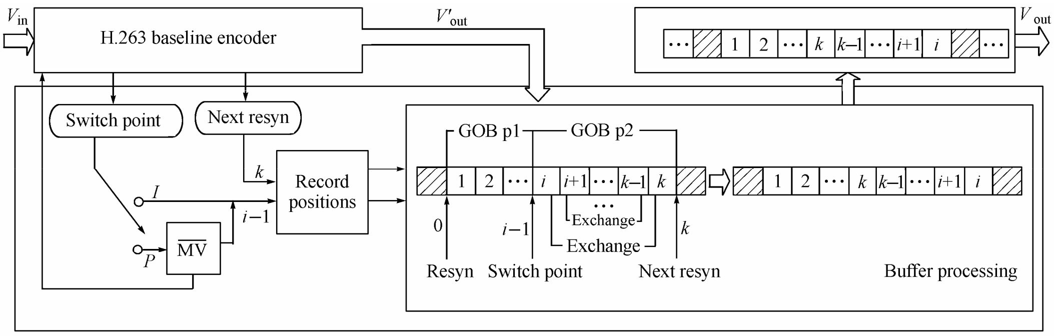

Fig.3 Schematic of the improved encoder

As mentioned above, I frames and P frames are basic backbones of H.263+ stream. For I frames, no modification should be specially considered in the improved mode because there is no MV data. But for P frame, MV data of current MB is obtained from its adjacent three MBs in baseline[5]. But in the improved encoder, the first motion vector of GOB p2 should be coded using a predictor value of 0 for both horizontal and vertical components to prevent the influence of the GOB p1, and the subsequent MBs' MVs are predicted using MVD as baseline does.

The block diagram of the improved decoder is shown in Fig.4. The emphasis of the decoder is on whether errors are detected in GOB p1 or not.

(1) If no error is detected in GOB p1 (illustrated in Fig.5), the decoder works like baseline until the first half MBs are completed and marks the end bit position of the MBs as $i-1$(switch point), and then searches the next synchronizer to find the last bit position k of the GOB. After that, the decoder reverses the bits from $i-1$ to k and then decodes these bits like baseline. And the discarded bits can be expressed as $E_1 = \sum_{j=i}^{k} P_e(j-i+1)$, where P_e denotes error probability of each bit position in GOBs.

(2) If error is detected in GOB p1 (illustrated in Fig.6), the decoder records the position j error occurred, searches the next synchronizer to get the GOB last bit position k, and then reverses the bits from j to k bit by bit. Then the decoder will process the data stream before the bit j as usual and later the successive MB number will jump to the first MB number of GOB p2 directly, then the decoder will decode the reversed bits, as shown in Fig.6. The discarded bits can be expressed as

$$E_2 = \sum_{j=1}^{i-1} P_e(i-j)$$

where P_e denotes error probability of each bit position in GOBs.

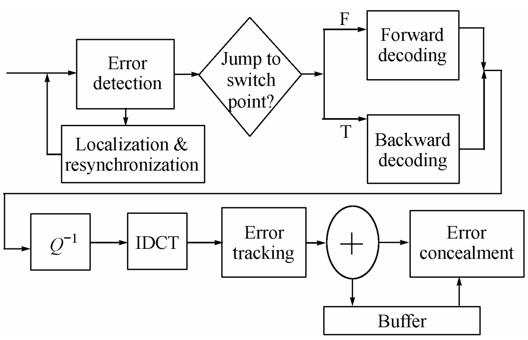

Fig.4 Diagram of the improved decoder

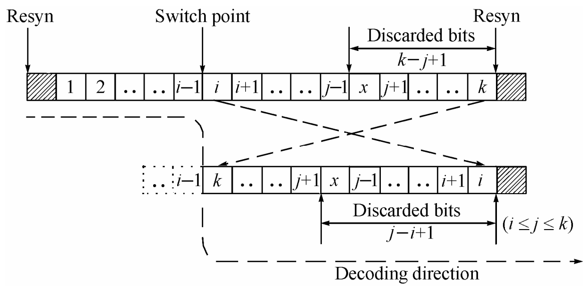

Fig.5 Decoding process when no error is detected in GOB p1

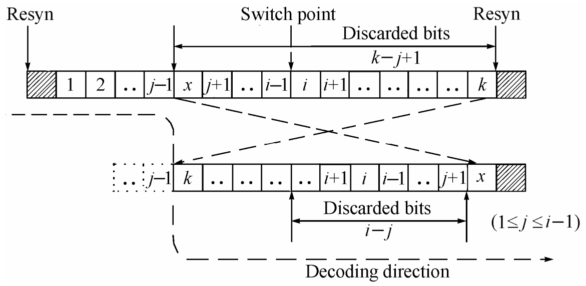

Fig.6 Decoding process when error is detected at bit position j

IV. Error Detection and Localization

Errors can be detected by codec syntax and intrinsic characteristic of video stream[4], such as large intensity jump coinciding with MB boundaries or block boundaries, inconsistent codes, invalid motion vectors, etc. With the assumption that video signals are smooth, we can detect lots of errors to improve the codec performance statistically.

According to syntax of H.263+, the following errors can be detected:

(1) Invalid VLC code, i.e. the code can not be found in H.263 standard VLC table;

(2) Motion vectors out of the range;

(3) DCT coefficients out of the range;

(4) Number of DCT coefficients in one block more than 64;

(5) Quantization step parameter out of range.

Generally, reconstructed MBs based on wrong information have much distortion than the concealed ones. So we should localize the beginning position of the corrupted MBs, which is called error localization. Considering the characteristics of natural video signals, in which most neighboring pixels do not vary greatly, an efficient error localization scheme is proposed by the following steps. Let MB_i represent the MB being checked and MB_{i+1} represent the immediately successive one.

(1) If MB_i locates at the left edge of a GOB, namely the first MB after synchronizer, intra-detection is used as shown in Fig.7. Calculate the absolute sum of difference between the pixels of the two adjacent rows and columns of the blocks in this MB. If the absolute sum is greater than a predetermined threshold, the MB is declared to be damaged, all the MBs in the GOB are considered to be corrupted. Otherwise replacing MB_i with MB_{i+1}.

(2) If MB_i does not locate at the left edge of a GOB, inter-detection is used as shown in Fig.8. Calculate the absolute sum of difference between the pixels of the left column of the current MB and the right column of its previous adjacent MB. If the absolute sum is greater than a predetermined threshold, the MB_i is declared to be damaged. Otherwise, replacing MB_i with MB_{i+1}, go back to (2).

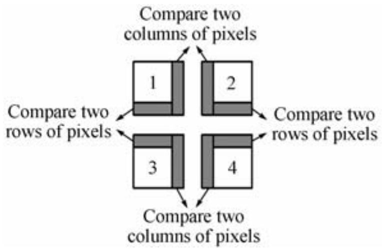

Fig.7 Intra-detection

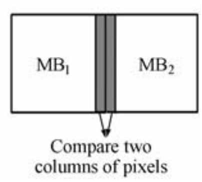

Fig.8 Inter-detection

V. Analysis and Explanation

In the improved decoder, the number of discarded bits because of transmission error can be expressed as

$$\begin{aligned} E &= E_1 + E_2 \\ &= \sum\nolimits_{j=1}^{i-1} P_e(i-j) + \sum\nolimits_{j=i}^{k} P_e(j-i+1) \\ &= \frac{k^2+3k}{2} P_e + (i^2 - ik - 2i + 1) P_e \end{aligned}$$

where P_e denotes error probability of each position in GOBs.

To evaluate the minimum value of E, let the derivative of E equal to zero and then we can get $dE = (2i - k - 2)P_e$. and then $dE = 0 \Rightarrow i = k/2 + 1$, which means that the minimum of E occurs when i locates at the middle of a GOB. That is to say, the minimum of discarded bits occurs when i locates at the middle of a GOB. For baseline, the switch point i can be assumed at the start of a GOB, namely i=0. So the improved scheme is much robust than baseline. When $i = k/2 + 1$, which is the optimized position of switch point, the number of lost data E can be expressed in terms of the error position j

$$E = \begin{cases} k/2 - j + 1, & 1 \le j \le k/2 \\ j - k/2, & k/2 + 1 \le j \le k \end{cases}$$

where k denotes the number of bits in the GOB shown in Fig.9. There are 29 bits as a GOB header, which is used as the resynchronization marker.

Fig.9 shows that different error position i has different influence on the performance of the decoder. If the errors occur near the optimized switch point $i = k/2 + 1$ in a GOB, few bits are discarded and almost all the data stream bits can be decoded.

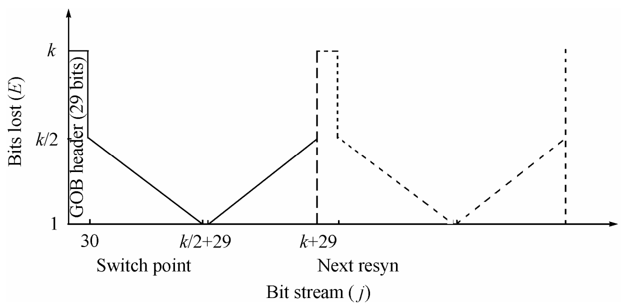

Fig.9 E expressed in terms of j, that is, $f(E, j)$ when $i = k/2+1$

VI. Experimental Results

The average Peak Signal to Noise Ratio (PSNR) has been widely adopted to describe the distortion. For error-free transmission, the PSNR reproduced at the receiver is given by[7]

$$\mathrm{PSNR}_{\mathrm{correct}}(t) = 10\log_{10}(255^2/D_{\mathrm{sc}}(t))$$

where $D_{\mathrm{sc}}(t)$ is Source Coding (SC) distortion given by $D_{\mathrm{sc}}(t) = U^{-1}\sum_{u=1}^{U}[o_u(t) - r_u(t)]^2$, and U is the total number of samples in the picture at time t, while $o_u(t)$ and $r_u(t)$ are the u-th amplitudes of the original and reconstructed pictures respectively.

Owing to transmission error, an additional distortion $D_{\mathrm{CH}}(t)$ occurs at the decoder. Thus the overall distortion (i.e. after error concealment) resulted is $D_0(t) = D_{\mathrm{SC}}(t) + D_{\mathrm{CH}}(t)$, and the PSNR at the decoder can be expressed as: $\mathrm{PSNR}_{\mathrm{lost}}(t) = 10\log_{10}(255^2/D_0(t))$. To measure the improvement of video degradation $\Delta\mathrm{PSNR}(t) = \mathrm{PSNR}_{\mathrm{improvement}}(t) - \mathrm{PSNR}_{\mathrm{baseline}}(t)$ is defined, where $\mathrm{PSNR}_{\mathrm{improvement}}(t)$ and $\mathrm{PSNR}_{\mathrm{baseline}}(t)$ are the PSNR of the improved scheme and baseline respectively.

The ITU-T error pattern files for WCDMA[8] are used to simulate the random errors of wireless channels. The encoded streams from baseline and improved encoders are disturbed respectively by the error pattern files W1~W6. The statistic results are listed in Tab.1 and the information sources are the standard video test sequences Foreman, Carphone and Grandma.

The reconstructed images from baseline and the improved codec in the same error patterns are shown in Figs.10~12 to observe the performance improvement subjectively.

VII. Conclusions

Error resilience has already been an important issue of video transmission over error-prone channels. Based on H.263+, by reversing part code stream sequences of each GOB, an error resilient coding scheme is studied to improve robustness without additional bandwidth burden. The improved method is analyzed theoretically and simulated using the ITU models. Error pattern files are employed to simulate WCDMA channels to check out the performance improvement. Simulation results show that both subjective and objective qualities of the reconstructed images are improved remarkably. The mean PSNR is increased by 0.5dB, and the highest increment is 2dB. The scheme along with suitable error concealment can be employed to transfer video clips in real-time efficiently through noisy channels.

Fig.10 Reconstructed images of the 25th, 60th and 105th (of 133) frames of "Foreman" corrupted by W1. The uppers are the outputs of baseline decoder and the lowers are the corresponding results of the improved decoders

Fig.11 Reconstructed images of the 4th, 21th and 60th (of 127) frames of "Carphone" corrupted by W3. The uppers are the outputs of baseline decoder and the lowers are the corresponding results of the improved decoders

Fig.12 Reconstructed images of the 40th, 100th and 165th (of 289) frames of "Grandma" corrupted by W5. The uppers are the outputs of baseline decoder and the lowers are the corresponding results of the improved decoders

Tab.1 The PSNR of H.263+ baseline and the improved scheme

	Foreman			Carphone			Grandma		
	Baseline	Improved	ΔPSNR	Baseline	Improved	ΔPSNR	Baseline	Improved	ΔPSNR
Error free	30.46	30.46	NA	31.39	31.39	NA	32.03	32.03	NA
W1	22.74	22.90	0.16	22.71	23.47	0.76	25.73	26.39	0.66
W2	28.24	28.70	0.46	29.11	29.51	0.40	31.68	31.73	0.05
W3	24.97	25.79	0.82	26.21	27.03	0.82	27.95	28.78	0.83
W4	29.37	293.62	0.25	29.93	30.34	0.41	31.34	31.59	0.25
W5	23.87	24.69	0.82	24.26	24.82	0.56	26.93	27.63	0.70
W6	28.93	29.26	0.33	30.43	30.60	0.17	31.25	31.51	0.26
Average	26.35	26.83	0.48	27.11	27.63	0.52	29.15	29.61	0.46

Note: W1: WCDMA-005Hz-3; W2: WCDMA-005Hz-4; W3: WCDMA-070Hz-3; W4: WCDMA-070Hz-4; W5: WCDMA-211Hz-3; W6: WCDMA-211Hz-4

References

[1] G. Y. Hong, B. Fong, A. C. M. Fong. Error localization for robust video transmission. *IEEE Trans. on Consumer Electronics*, **48**(2002)3, 463−469.

[2] Tu-Chih Wang, Hung-Chi Fang, Liang-Gee Chen. Low-delay and error-robust wireless video transmission for video communications. *IEEE Trans. on Circuits and Systems for Video Technology*, **12**(2002) 12, 1049−1058.

[3] Bernd Girod, Niko Farber. Feedback-based error control for mobile video transmission. *Proc. of the IEEE Special Issue on Video for Mobile Multimedia*, **87**(1999) 10, 1707−1723.

[4] Shaoshuai Gao, Guofang Tu. Robust H.263+ video transmission using partial backward decodable bit stream (PBDBS). *IEEE Trans. on Circuits and Systems for Video Technology*, **13**(2003)2, 182−187.

[5] International Telecommunication Union Draft ITU-T Recommendation H.263. Video coding for low bit rate communication. ITU-T, 1996.

[6] J. Wen, J. D. Villasenor. Reversible variable length codes for efficient and robust image and video coding. Proc. of Data Compression Conf., Snowbird, UT, 1998, 471−480.

[7] Christian Dubuc, Daniel Boudreau, Franois Patenaude. The design and simulated performance of a mobile video telephony application for satellite third-generation wireless systems. *IEEE Trans. on Multimedia*, **3**(2001)4, 424−431.

[8] Ericsson, Lund, Sweden, Standardization Sector of ITU, WCDMA Error Patterns at 64 kb/s. Proc. ITU-T Study Group 16. Multimedia Terminals and Systems Expert Group, 1998.