OPTIMIZED NESTED PROTECTION FOR VIDEO REGION OF INTEREST WITH RAPTOR CODES

Zhengyi Luo[1,2], *Li Song*[1,2,3], *Shibao Zheng*[1,2] *and Nam Ling*[3]
[1]Shanghai Digital Media Processing and Transmission Key Lab
[2]Shanghai Jiao Tong University, Shanghai, China
[3]Santa Clara University, Santa Clara, USA

ABSTRACT

Due to the best effort feature of many existing transmission channels, video streams often suffer from inevitable transmission errors. In this paper, we propose a scheme of robust video transmission based on the state-of-the-art Raptor codes, whose applications are in full swing now. And considering Region of Interest (ROI) often draws much attention in images, the scheme adopts a nested protection framework to show partialities to ROI areas for better protection. Different from many existing Raptor codes based UEP methods, our scheme is developed based on the easy-to-use standardized Raptor codes. Experimental results show that significant robustness can be obtained for the video streams, especially for the ROI areas.

Index Terms**—** UEP, Raptor codes, video, ROI

1. INTRODUCTION

With the development of multimedia technologies, applications via video transmission have played an important role in modern life. However, existing transmission channels, such as the Internet and wireless networks, usually provide only best effort services. Therefore, ways of robust video transmission are always desired by many applications.

As far as the transmission level is concerned, a common way of robust video transmission is Automatic Repeat Request (ARQ), which makes the receiver send retransmission requests to the sender if video packets are lost. Obviously ARQ necessitates feedback channels and introduces extra delay, and is unrealistic for many applications such as the broadcasting services.

Besides ARQ, another way of robust transmission is Forward Error Correction (FEC), which provides robustness based on channel codes. Different from ARQ, FEC requires no feedback channels but redundancy symbols to combat the transmission errors, and thus is quite popular in many applications such as live video streaming.

Frequently, Reed-Solomon (RS) codes are used as the FEC codes to provide robust video transmission. But they have high computational complexity and are practical only at small block lengths. Recently, Raptor codes [1], a new kind of fountain channel codes [2], have been developed. Compared to conventional channel codes, e.g. RS codes, Raptor codes have low computational complexity and are still practical at large block lengths. And systematic Raptor codes have been standardized and adopted as FEC codes in latest multimedia standards [3][4]. Now in view of the superiority of Raptor codes, applications of Raptor codes in the multimedia field are in full swing [5][6].

Recently Raptor codes have also been utilized to provide robust video transmission via Unequal Error Protection (UEP). Vukobratovic et al. [7] proposed an UEP method for scalable video multicast based on expanding window fountain codes, which incorporated a window selection procedure in fountain coding. Cataldi et al. [8] applied UEP based on sliding window Raptor codes to scalable video broadcasting, where Raptor codes were improved by virtually extending the block length with sliding windows. Both of these methods adapt to UEP frameworks by modifying the structure of Raptor codes, and may incur extra difficulties in practice compared with the standardized Raptor codes. In this paper, we take a different routine and propose a content based UEP optimization scheme based on the standardized systematic Raptor codes. This scheme provides different degree of protection to packets of different types at the Group of Pictures (GOP) level, and adopts a nested protection framework considering that Region of Interest (ROI) often draws much attention and should be protected more than others. Compared with Hellge et al. [9], here packet classification is introduced and the expected distortion impacts are minimized for better protection. Experimental results show that our scheme can provide significant robustness for video transmission, especially for the ROI areas.

The remainder of this paper is organized as follows. Section 2 introduces Raptor codes briefly. Section 3 describes the adopted protection framework. Section 4 formulates the UEP problem. Experimental results validating the effectiveness of the scheme are shown in Section 5. And Section 6 draws the conclusion.

2. RAPTOR CODES

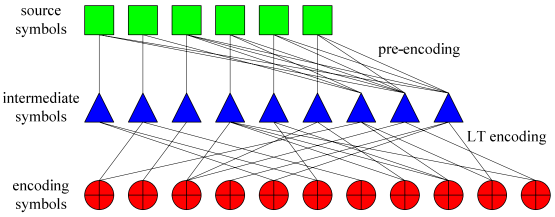


Fig. 1. Illustration of the encoding of Raptor codes.

Fountain codes [2] are a class of channel codes, whose encoders can generate as many encoding symbols as necessary from a given number of source symbols just as fountains spray water drops. At the decoder side, the source symbols can be recovered as long as enough encoding symbols are received.

The first kind of practical fountain codes are Luby Transform (LT) codes [10], whose encoding can be divided into two steps. First, the degree d of an encoding symbol is chosen from a degree distribution, e.g. the Robust Soliton Distribution [10]. Then d distinct source symbols are chosen uniformly and randomly and exclusive-ORed to produce the encoding symbol.

Raptor codes are a new kind of fountain codes, which generally build upon LT codes and LDPC codes [11]. The encoding of Raptor codes can be illustrated in Figure 1, where the source symbols are first pre-encoded into the intermediate symbols, which are then LT encoded into the final encoding symbols. At the decoder side, the source symbols can be recovered from any subset of encoding symbols only slightly more in number than the number of source symbols. Raptor codes have the advantage of linear time encoding and decoding, and especially benefit large block lengths. Now in light of the preference of systematic codes, systematic Raptor codes have been developed and adopted in standards like 3GPP [3]. And due to the good performance Raptor codes have become an excellent option for robust multimedia transmission [5][6].

3. ADOPTED PROTECTION FRAMEWORK

It is well known that different video packets, such as intra (I), predictive (P) or bi-directional (B) packets, have different distortion impacts, and UEP for different packet types usually can provide enhanced robustness for video transmission. Hence, we may design the Raptor codes based UEP framework in accordance with packet types at the GOP level at first. Further, research on Human Visual System (HVS) reveals that people generally pay more attention to the ROI areas in images, which motivates us to design the protection framework in a nested manner, so that ROI areas have better error resilience.

Based on the above two key points, we firstly classify the video packets of a GOP into N=3 classes — class 1 for I packets, class 2 for P packets and class 3 for B packets, and then further divide each class i into subclass iF for ROI packets and subclass iB for background packets. Our scheme processes each class of packets separately but in similar procedures. So the nested protection framework is described with one representative class of packets in the following.

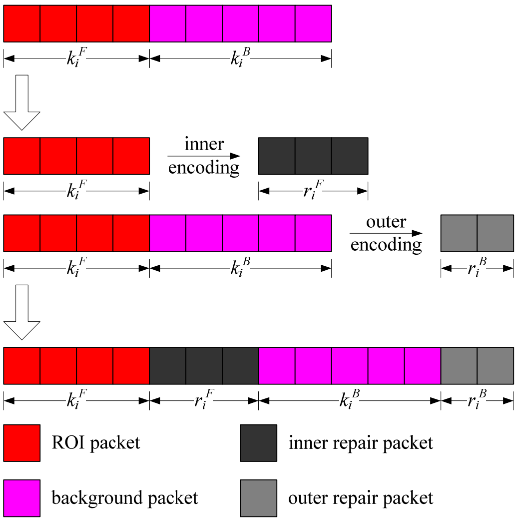


Fig. 2. Illustration of the adopted protection framework.

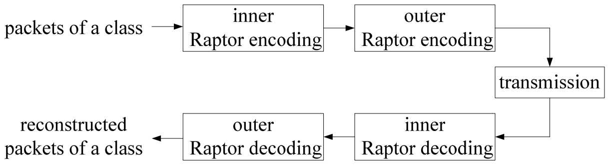


Fig. 3. Illustration of the Raptor encoding and decoding procedures of each class.

In the context of Raptor protection here, every video packet is regarded as a Raptor symbol. The nested framework involves inner and outer Raptor encoding, where the former protects part of packets while the latter protects all packets. As illustrated in Figure 2, each class i is composed of subclass iF and subclass iB, which contain respectively ROI packets and background packets of class i. As ROI packets often have higher distortion impacts than background packets, we first perform the inner Raptor encoding, which takes the ROI packets of class i as source packets and produces the inner repair packets. Since systematic Raptor codes are adopted in our framework, after inner encoding the ROI packets from subclass iF remain and we can continue to perform the outer Raptor encoding, which takes both the ROI and the background packets as source packets and produces the outer repair packets. In this way ROI packets can be recovered by both the inner and the outer repair packets, while background packets can be recovered only by outer repair packets.

The Raptor encoding and decoding procedures of each class can be shown in Figure 3. Note that at the receiver side, firstly the inner Raptor decoding is performed for the ROI packets. If the inner Raptor decoding succeeds, the background packets may be recovered more easily as the outer Raptor decoding has more encoding symbols, which have been partly recovered by the inner decoding. But even if the inner Raptor decoding fails, ROI packets may still be recovered by the outer Raptor decoding. Therefore, the inner and outer Raptor encoding gives ROI packets a higher probability of recovery than background packets. And the nested protection framework not only enables robust video transmission but also provides higher priorities to ROI packets.

4. PROBLEM FORMULATION AND OPTIMIZATION

Our scheme provides a robust video protection framework, but it still has to be properly configured for higher efficiency. To this end, we formulate the protection based on the above nested framework as a constrained optimization problem, and determine its optimal configuration in terms of the source and channel conditions.

For description purposes, we first clarify some adopted notations. At the random packet loss rate p_l, the probability that m out of n packets are received is

$$P_b\left(n,m\right)=C_n^m\left(1-p_l\right)^m p_l^{\,n-m}. \tag{1}$$

Assume the standardized systematic Raptor codes in 3GPP [3] are used in this paper. $P_f\left(m,k\right)$ represents the failure probability of the standardized Raptor code with k source symbols if m symbols are received, and can be modeled by [12]

$$P_f\left(m,k\right)=\begin{cases}1 & m<k\\ 0.85\times0.567^{m-k} & m\geq k\end{cases}. \tag{2}$$

As illustrated in Figure 2, let k_i^F and k_i^B denote respectively the number of ROI packets and background packets of class i, and let r_i^F and r_i^B denote respectively the number of the inner and outer repair packets of class i. Assume m_a out of k_i^F ROI packets, m_b out of r_i^F inner repair packets and m_c out of $k_i^B+r_i^B$ background and outer repair packets are received.

As the ROI packets fail to recover when both the inner and outer Raptor decoding fails, the probability that a ROI packet of subclass iF gets lost finally can be approximated by

$$\begin{aligned}P_i^F\approx&\sum_{m_a=0}^{k_i^F}\sum_{m_b=0}^{r_i^F}\sum_{m_c=0}^{k_i^B+r_i^B}P_b\left(k_i^F,m_a\right)\cdot P_b\left(r_i^F,m_b\right)\\&\cdot P_b\left(k_i^B+r_i^B,m_c\right)\cdot P_f\left(m_a+m_b,k_i^F\right)\\&\cdot P_f\left(m_a+m_c,k_i^F+k_i^B\right)\cdot\frac{C_{k_i^F-1}^{m_a}}{C_{k_i^F}^{m_a}}\end{aligned}, \tag{3}$$

where $C_{k_i^F-1}^{m_a}\Big/C_{k_i^F}^{m_a}$ is the probability that the video packet is not included in the m_a packets due to systematic Raptor codes. The background packets fail to recover when the outer Raptor decoding fails. As more encoding symbols will be available for the outer Raptor decoding if the inner Raptor decoding succeeds, the probability that a background packet of subclass iB gets lost finally can be approximated by

$$\begin{aligned}P_i^B\approx&\sum_{m_a=0}^{k_i^F}\sum_{m_b=0}^{r_i^F}\sum_{m_c=0}^{k_i^B+r_i^B}P_b\left(k_i^F,m_a\right)\cdot P_b\left(r_i^F,m_b\right)\\&\cdot P_b\left(k_i^B+r_i^B,m_c\right)\cdot\left\{\left[1-P_f\left(m_a+m_b,k_i^F\right)\right]\right.\\&\cdot P_f\left(k_i^F+m_c,k_i^F+k_i^B\right)+P_f\left(m_a+m_b,k_i^F\right)\\&\left.\cdot P_f\left(m_a+m_c,k_i^F+k_i^B\right)\right\}\cdot\frac{C_{k_i^B+r_i^B-1}^{m_c}}{C_{k_i^B+r_i^B}^{m_c}}\end{aligned}, \tag{4}$$

where $C_{k_i^B+r_i^B-1}^{m_c}\Big/C_{k_i^B+r_i^B}^{m_c}$ is the probability that the video packet is not included in the m_c packets.

Let $D_i^F\left(j\right)$ denote the distortion impact due to the loss of the j th packet of subclass iF, and $D_i^B\left(j\right)$ the distortion impact due to the loss of the j th packet of subclass iB. Assume the distortion impacts of video packets do not correlate with each other for simplicity. Then the expected distortion impacts of all the video packets due to packet loss can be approximated by

$$\begin{aligned}E\left[D\right]&=\sum_{i=1}^{N}\left[\sum_{j=1}^{k_i^F}P_i^F\cdot D_i^F\left(j\right)+\sum_{j=1}^{k_i^B}P_i^B\cdot D_i^B\left(j\right)\right]\\&=\sum_{i=1}^{N}\left[P_i^F\cdot\sum_{j=1}^{k_i^F}D_i^F\left(j\right)+P_i^B\cdot\sum_{j=1}^{k_i^B}D_i^B\left(j\right)\right]\end{aligned}, \tag{5}$$

where packets of $N=3$ (I, P and B) classes are all considered. For higher video quality, we can try to minimize the expected distortion impacts by optimizing the distribution of repair packets. Specifically, suppose R repair packets are available at the given FEC ratio, then the problem can be formulated as

$$\begin{aligned}&\left\{\widehat{r_1^F},\widehat{r_1^B},\cdots,\widehat{r_N^F},\widehat{r_N^B}\right\}=\underset{r_1^F,r_1^B,\cdots,r_N^F,r_N^B}{\arg\min}\;E\left[D\right]\\&s.t.\begin{cases}\sum_{i=1}^{N}\left(r_i^F+r_i^B\right)\leq R\\ r_i^F\geq0\;\left(1\leq i\leq N\right)\\ r_i^B\geq0\;\left(1\leq i\leq N\right)\\ \dfrac{r_i^F+r_i^B}{k_i^F+k_i^B}\geq\dfrac{r_{i+1}^F+r_{i+1}^B}{k_{i+1}^F+k_{i+1}^B}\;\left(1\leq i\leq N-1\right)\end{cases}\end{aligned}, \tag{6}$$

where the last constraint is the descending priority constraint accounting for the differences of frame types.


(a) (b)

Fig. 4. ROI areas of (a) Raven and (b) Sheriff.

Through the penalty method, the problem can be translated into minimizing

$$E[D]^* = E[D] + \lambda_1 \cdot \max\left[0, \sum_{i=1}^{N}\left(r_i^F + r_i^B\right) - R\right] + \sum_{i=1}^{N-1}\lambda_{1+i}\cdot\max\left[0, \left(k_i^F+k_i^B\right)\left(r_{i+1}^F+r_{i+1}^B\right)-\left(k_{i+1}^F+k_{i+1}^B\right)\left(r_i^F+r_i^B\right)\right], \quad (7)$$

where all λ_i are penalty coefficients, and solved by the heuristic algorithms such as the simulated annealing algorithm used in this paper.

5. EXPERIMENTS

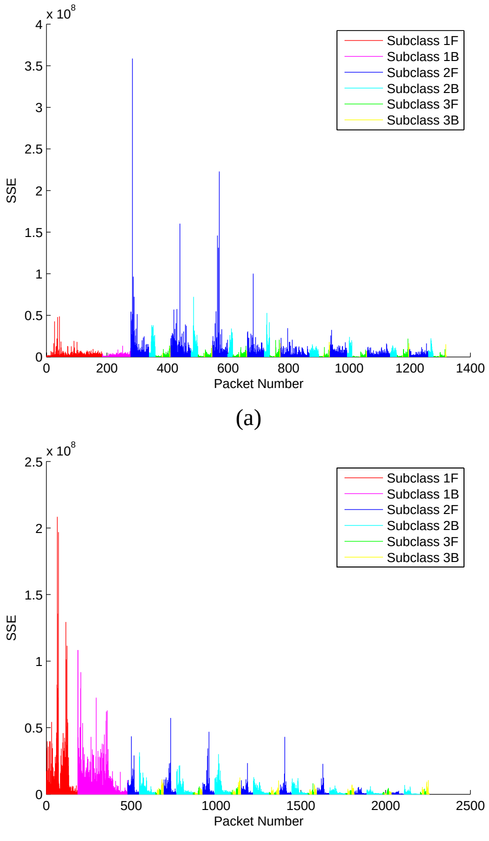
The proposed scheme was evaluated with the representative 1280×720 4:2:0 sequences Raven and Sheriff. Since the scheme is configured and applied GOP by GOP, here we present the full parameter configuration and experimental results for the first GOP of the test sequences, whose ROI areas are indicated in Figure 4 and are supposed not to change in the GOP. It should be noted that our scheme does not depend on the setting of ROI areas, and the above setting is used here only to facilitate comparison in the following. In fact, the scheme can be applied freely with other setting of ROI areas as needed.

We used the JM 14.2 software [13] to obtain H.264/AVC video packets of 150 bytes, where Flexible Macroblock Ordering (FMO) type 2 was enabled for ROI encoding. GOPs of IBBPBBP… structure including 25 frames were considered. QP=28 and CABAC was adopted during compression.

It was supposed that video packets were lost independently, and 10% FEC ratio ($R\Big/\left[\sum_{i=1}^{N}\left(k_i^F+k_i^B\right)+R\right]=10\%$) was allowed to mitigate transmission errors. In the experiments, the error concealment algorithm for I frames was weighted intra interpolation [14], and that for P and B frames was direct copy.

The distortion impacts of video packets were measured by Sum of Squared Errors (SSE). Let M, W and H denote respectively the frame number in a GOP, the frame width and the frame height. Then the distortion impact of a packet can be expressed as

$$D=\sum_{m=1}^{M}\sum_{x=1}^{W}\sum_{y=1}^{H}\left[f(m,x,y)-\hat{f}(m,x,y)\right]^2, \quad (8)$$



(a)

(b)

Fig. 5. Classification and distortion impacts of the video packets for the first GOP of (a) Raven and (b) Sheriff.

where $f(m,x,y)$ and $\hat{f}(m,x,y)$ are respectively the pixels of the correctly decoded sequence and the decoded sequence with that packet error concealed. The classification results along with the distortion impacts of the video packets for the test sequences are shown in Figure 5. It can be seen that the distortion impact has quite a high order of value, so the penalty coefficients were set to 10^{10} in the penalty method. Based on such classification, distribution of the repair packets was optimized by the simulated annealing algorithm. At different packet loss rates, after a certain number of iterations the obtained configuration parameters of the proposed scheme are shown in Table 1, where the number of source packets for the test sequences is also shown.

As mentioned before, Raptor codes have much superiority over conventional channel codes. In addition, other main Raptor codes based UEP methods involve more or less modifying the standardized Raptor codes, while our scheme builds upon the easy-to-use standardized Raptor codes. Therefore, based on the configuration parameters, the proposed scheme was compared with the standardized Raptor codes based Equal Error Protection (EEP) case, which was implemented by protecting the video packets

Table 1. Configuration parameters for the first GOP of the test sequences

Sequence	Packet Loss Rate	Parameter											
		k_1^F	r_1^F	k_1^B	r_1^B	k_2^F	r_2^F	k_2^B	r_2^B	k_3^F	r_3^F	k_3^B	r_3^B
Raven	0.1	184	0	94	44	526	0	167	103	298	0	51	0
	0.125		43		0		104		0		0		0
	0.15		43		0		104		0		0		0
	0.2		0		147		0		0		0		0
	0.25		0		147		0		0		0		0
	0.3		0		147		0		0		0		0
	0.35		147		0		0		0		0		0
	0.4		147		0		0		0		0		0
Sheriff	0.1	185	0	294	75	572	0	764	176	183	0	259	0
	0.125		0		119		132		0		0		0
	0.15		0		119		132		0		0		0
	0.2		0		248		1		2		0		0
	0.25		0		251		0		0		0		0
	0.3		0		251		0		0		0		0
	0.35		245		0		4		2		0		0
	0.4		251		0		0		0		0		0

equally in a block considering the good performance of Raptor codes at large block lengths. At different packet loss rates, the measured PSNR for the ROI areas and the whole sequences is shown in Figure 6, where all the reported PSNR is averaged over 100 simulations. It can be seen that compared with EEP, obvious gains can be obtained for both the ROI areas and the whole sequences. This is because repair packets are distributed according to the distortion impacts of video packets in our scheme. What is more, we can also see that ROI areas may have higher PSNR improvement than the whole sequences. This is because we adopt a nested protection framework, which gives higher priorities of error protection and repair packet distribution to ROI areas. The reason why Raven does not have as much ROI improvement as Sheriff is that its ROI areas account for most of the image areas and the resulting video packets. But this can be improved by adjusting the size of the ROI areas.

Figure 7 shows the PSNR comparison for every frame under a same loss pattern of the packet loss rate 0.15. Figure 8 and Figure 9 show respectively part of the 14th reconstructed frame of Raven and the 16th reconstructed frame of Sheriff in this case. We can also see that our scheme performs better.

6. CONCLUSION

In this paper, we design a scheme of robust video transmission based on the standardized Raptor codes. We adopt a nested protection framework so that ROI packets can have higher priorities of error protection during transmission. Experimental results show that significant robustness can be obtained for the video streams, especially for the ROI areas.

7. ACKNOWLEDEMENT

This work was supported by the national 863 Program (2012AA011703), the china MIIT Program (2010ZX03004-003), the NSF grants (60902020, 61102098, and 60902073), STCSM (12DZ2272600) and the 111 project.

8. REFERENCES

[1] A. Shokrollahi, “Raptor codes,” IEEE Transactions on Information Theory, vol. 52, no. 6, pp. 2551-2567, June 2006.

[2] D. J. C. MacKay, “Fountain codes,” IEE Proceedings - Communications, vol. 152, no. 6, pp. 1062-1068, December 2005.

[3] 3GPP, “Multimedia broadcast/multicast service (MBMS); protocols and codecs,” 3GPP TS 26.346 v10.0.0, March 2011.

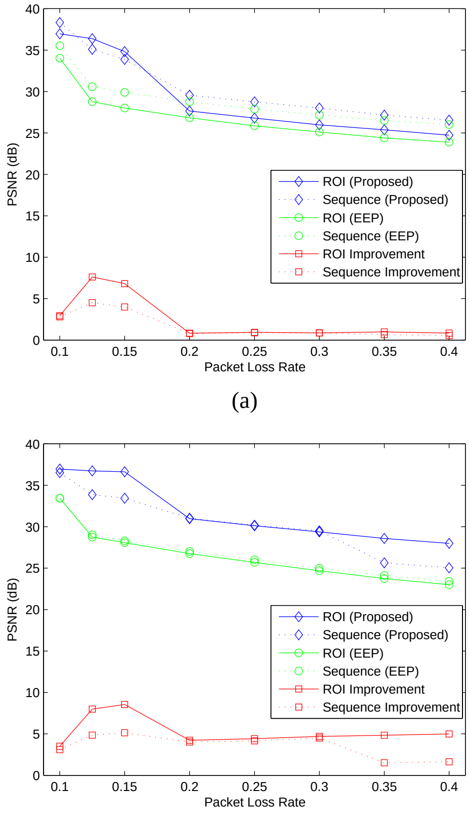
[4] M. Luby, A. Shokrollahi, M. Watson and T. Stockhammer, “Raptor forward error correction scheme for object delivery,” RFC 5053, October 2007.

[5] T. Stockhammer and A. C. Begen, “Response for MMT EE #9: application layer FEC in the IETF,” MPEG M21546, October 2011.

[6] T. Stockhammer, “IETF standards on FEC,” MPEG M22721, November 2011.

[7] D. Vukobratovic, V. Stankovic, D. Sejdinovic, L. Stankovic and Z. Xiong, “Scalable video multicast using expanding window fountain codes,” IEEE Transactions on Multimedia, vol. 11, no. 6, pp. 1094-1104, October 2009.

[8] P. Cataldi, M. Grangetto, T. Tillo, E. Magli and G. Olmo, “Sliding-window Raptor codes for efficient scalable wireless video broadcasting with unequal loss



(a)

(b)

Fig. 6. PSNR comparison for the first GOP of (a) Raven and (b) Sheriff under random packet loss.

protection," IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1491-1503, June 2010.

[9] C. Hellge, D. Gomez-Barquero, T. Schierl and T. Wiegand, "Layer-aware forward error correction for mobile broadcast of layered media," IEEE Transactions on Multimedia, vol. 13, no. 3, pp. 551-562, June 2011.

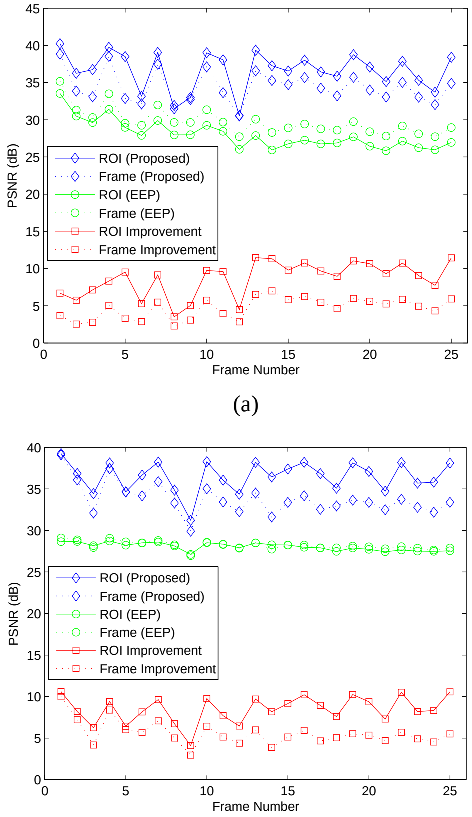
[10] M. Luby, "LT codes," Proceedings of the IEEE Symposium on Foundations of Computer Science, pp. 271-280, Vancouver, November 2002.

[11] R. G. Gallager, Low-Density Parity-Check Codes. Cambridge, Massachusetts: MIT Press, 1963.

[12] M. Luby, T. Gasiba, T. Stockhammer and M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," IEEE Transactions on Broadcasting, vol. 53, no. 1, pp. 235-246, March 2007.

[13] "JM 14.2 reference software," Available: http://iphome.hhi.de/suehring/tml/download/.

[14] Y.-K. Wang, M. M. Hannuksela, V. Varsa, A. Hourunranta and M. Gabbouj, "The error concealment feature in the H.26L test model," Proceedings of the IEEE International Conference on Image Processing, pp. II-729-II-732, Rochester, New York, September 2002.



(a)

(b)

Fig. 7. PSNR comparison for the first GOP of (a) Raven and (b) Sheriff under a same loss pattern of the packet loss rate 0.15.

(a) (b)

Fig. 8. Part of the 14th reconstructed frame of Raven for (a) EEP and (b) our scheme under a same loss pattern of the packet loss rate 0.15.

(a) (b)

Fig. 9. Part of the 16th reconstructed frame of Sheriff for (a) EEP and (b) our scheme under a same loss pattern of the packet loss rate 0.15.