

# Low-Complexity Video Coding for Wireless Image Transmission in Capsule Endoscopy

Kenichi Takizawa  
National Institute of Information and  
Communications Technology (NICT)  
3-4 Hikrarino-oka  
Yokosuka, Japan  
takizawa@nict.go.jp

Ryu Miura  
National Institute of Information and  
Communications Technology (NICT)  
3-4 Hikrarino-oka  
Yokosuka, Japan  
ryu@nict.go.jp

## ABSTRACT

This paper presents a low-complexity video compression technique for wireless transmission in capsule endoscopy. The capsule endoscopy uses a capsule device to convey images taken in the digestive tract wirelessly to a body-worn device. Since the capsule device has small capacity on its battery, compression technique is crucial that is realized in less power consumption. The technique shown in this paper is theoretically based on Slepian-Wolf coding, in which side information available at transmitter side is treated as well as side information at its receiver side. Therefore, energy-hungry processes for achieving video compression including estimation of motion vectors are moved to the receiver side. Simulation results show that the encoding method provides compression rate close to its lower bound for an 8-bit quantized raw (Bayer) data by using a regular low-density parity check (LDPC) under additive white Gaussian noise (AWGN) channel.

## 1. INTRODUCTION

A capsule endoscopy has been viewed as a valuable medical diagnosis since it enables to observe the inside of small intestine without patients' pain through wireless video transmission [1]. Current capsule products transmit a couple of frames per second because of limitation on its available channel bandwidth. However, there has been a demand of the increase of the frame rate in order to provide more reliable diagnosis. To achieve a higher frame rate, a compression technique applicable into a capsule device, which has a small-capacity battery, is crucial.

On the compression for a capsule endoscopy, one candidate is the use of lossless JPEG [2], which is a combination of differential pulse coded modulation (DPCM) and entropy coding. Another candidate is lossy-JPEG [3], which is a combination of discrete cosine transform (DCT) and Huffman coding. However, these JPEG compression techniques requires large amount of memory or additional processing;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

PFT 2013, September 30-October 02

Copyright © 2013 ICST 978-1-936968-89-3

DOI 10.4108/icst.bodynets.2013.253712

Figure 1: A block diagram of a capsule endoscopy using DVC through BPSK transmission under AWGN channel.

the lossy JPEG needs a buffer memory to store the previous frame, and both lossless and lossy JPEGs employ entropy coding as additional coding. H.264, has been viewed as the most efficient coding for video images; but its encoding process includes energy-hungry computations such as motion estimation, DCT, and arithmetic coding. As a result, these coding methods are not applicable for a capsule device from a power-consumption viewpoint. Recently, several works have been done towards realizing a compression technique in low-power consumption [4, 5].

In this paper, another framework called distributed video coding (DVC) [6] is introduced as compression technique for the capsule device. This compression method is theoretically based on Slepian-Wolf coding [7]; side information available at its transmitter (encoder) side such as a correlation between successive two frames is utilized at the receiver (decoder) side. Based on this framework, it is achieved to shift energy-hungry process at the encoder to the decoder, which is a body-worn device available to utilize larger-capacity battery in capsule endoscopy. Performance evaluation is given from the viewpoint of rate distortion characteristics under AWGN channels. Also, a comparison of loss-less JPEG and lossy-JPEG is addressed from their rate distortion characteristics.

## 2. VIDEO CODING AND DECODING BASED ON DVC

### 2.1 Overview

A low-complexity video encoding scheme based on DVC is presented. A block diagram of the system is illustrated in Fig. 1. The video coding block, which is applied into raw data, is implemented with an error correction coding (ECC) such as regular LDPC codes. The codeword is then transmitted to a body-worn device, which is a receiver side, wirelessly. In this paper,  $\pi/2$ -shift binary phase shift keying (BPSK) transmission is used since this modulation scheme provides an envelope in a small variation of its amplitude. At the receiver side, after its transmitted raw data is recovered through video decoding which is achieved by decoding for the ECC applied in its transmitter side, RGB image data is obtained through demosaicing. The video decoding utilizes

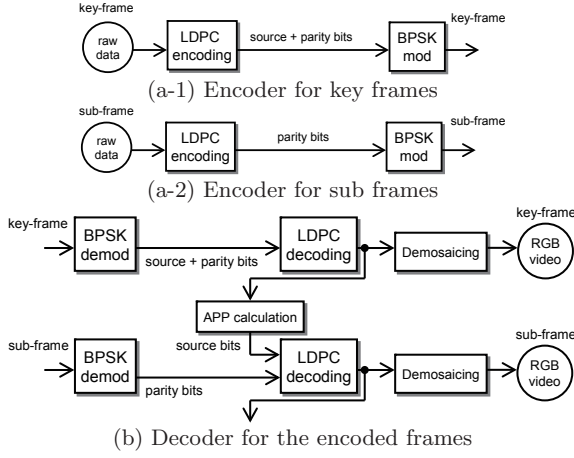


Figure 2: An illustration on the procedure of the proposed method on video encoding (a-1 and a-2) and decoding (b).

correlation between successive frames as side information in order to realize decoding based on Wyner-Ziv coding. Also, estimation on motion vector is introduced at the decoding process.

## 2.2 Encoding

Fig. 2 shows procedures of the presented video encoding and decoding methods. Here, a regular LDPC code is introduced as its ECC to achieve data compression. First, the original raw data is encoded at LDPC encoder at every frame, where the video frame is categorized into two different types; one is a key frame and the other is a sub frame as well as ordinary video compression methods. A sequence is consisted of a key frame that is its initial frame and multiple sub frames followed by the key frame. The number of sub frames is denoted by  $N_f$ . For the key frame, both source and parity data are transmitted; on the other hand, only parity data is conveyed for the sub frames in order to achieve video compression. When using an ECC with coding rate of  $R_c$ , achievable compression rate  $R$  is lower-bounded by the following equation.

$$R = \frac{1}{N_f} \left\{ \frac{1}{R_c} + (N_f - 1) \left( \frac{1}{R_c} - 1 \right) \right\} \quad (1)$$

Fig. 3 shows its compression rate  $R$  given by the Eq. (1). This lower-bound is not viewpoint from information theory that provides different lower-bound derived from the maximum conditional entropy between successive frames [7].

The detail of its encoding process is as follows. The amplitude of raw data is quantized in  $N_B$  bits per a pixel. The quantized bits in the  $i$  th frame image are denoted as  $b_{x,y,k}^i$ , where  $x \in \{1, \dots, x_{\max}\}$  and  $y \in \{1, \dots, y_{\max}\}$  respectively stand for the horizontal and vertical index of the pixel, and  $k \in \{1, \dots, N_B\}$  represents the bit-depth of the pixel. The frame image is also represented by  $\mathbf{B}^i = \{\mathbf{b}_{1,1}^i, \dots, \mathbf{b}_{x,y}^i, \dots, \mathbf{b}_{x_{\max},y_{\max}}^i\}$  where  $\mathbf{b}_{x,y}^i = \{b_{x,y,1}^i, \dots, b_{x,y,N_B}^i\}$ . This frame image is encoded at the ECC encoder in the following order:

$$\mathbf{B}_{\text{ecc}}^i = \{\mathbf{b}^i(0), \dots, \mathbf{b}^i(n), \dots, \mathbf{b}^i(x_{\max} \cdot y_{\max} - 1)\} \quad (2)$$

where  $\mathbf{b}^i(n) = \mathbf{b}_{x_n,y_n}^i$  and  $x_n = \text{rem}(n, x_{\max}) + 1$ , and

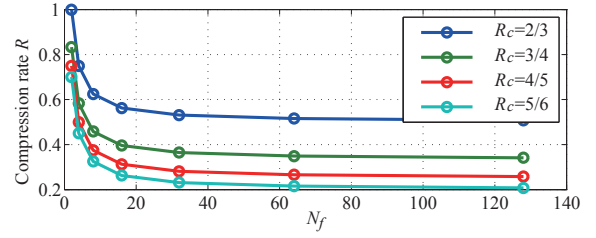


Figure 3: Achievable compression rate  $R$  given by the video encoding using an ECC.

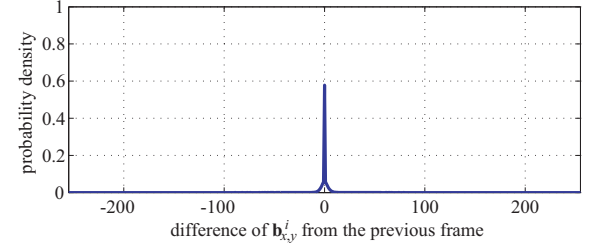


Figure 4: Probability distribution on the difference of pixel value  $\mathbf{b}_{x,y}^i$  from the previous frame.

$y_i = \text{rem}(\text{floor}(n, x_{\max}) + \text{rem}(n, y_{\max}), y_{\max}) + 1$ . Here,  $\text{rem}(n, m)$  represents the remainder of  $n$  divided by  $m$ , and  $\text{floor}(n, m)$  returns the largest integer less than  $n$  divided by  $m$ . The bit sequence  $\mathbf{B}_{\text{ecc}}^i$  is encoded by  $L_{\text{info}}$  bits which is the information length of an employed LDPC code. The number of codewords per a frame image is given by  $N_w = \lceil \frac{N_B \cdot x_{\max} \cdot y_{\max}}{L_{\text{ecc}}} \rceil$ . For a key frame, both its source and parity data, where its codeword is denoted as  $L_{\text{ecc}}$ , are transmitted for the frame image. On the other hand, for sub frames, only parity data with length of  $L_{\text{ecc}} - L_{\text{info}} = (\frac{1}{R_c} - 1) \cdot L_{\text{info}}$  bits is conveyed.

## 2.3 Decoding

Fig. 2 also shows the decoding procedure. For the key frame, ordinary decoding scheme for the employed LDPC is applied. On the other hand, since the sub frames has only parity bits at the receiver side, an estimated frame image is constructed from recovered image at the previous frame. This construction is achieved by employing side information that includes a correlation characteristic between successive image frames and estimates of motion vectors. As a priori information corresponding to its source data, log likelihood ratio (LLR) is computed on each bit for the sub frames. The a priori LLR for the  $i$  th frame denoted as  $L(b_{x,y,k}^i | \mathbf{m}_j, \mathbf{b}_{x,y}^{i-1})$  is given by

$$L(b_{x,y,k}^i | \mathbf{m}_j, \mathbf{b}_{x,y}^{i-1}) = \log \frac{\sum_{b_{x,y,k}^i=1} \Pr(\mathbf{b}_{x,y}^i | \mathbf{b}_{x-m_j,x-y-m_j,y}^{i-1}) \Pr(\mathbf{b}_{x-m_j,x-y-m_j,y}^{i-1})}{\sum_{b_{x,y,k}^i=0} \Pr(\mathbf{b}_{x,y}^i | \mathbf{b}_{x-m_j,x-y-m_j,y}^{i-1}) \Pr(\mathbf{b}_{x-m_j,x-y-m_j,y}^{i-1})} \quad (3)$$

where  $\Pr(\mathbf{b}_{x,y}^i | \mathbf{b}_{x,y}^{i-1})$  is a transition probability on a pixel value from  $\mathbf{b}_{x,y}^{i-1}$  at the  $i-1$  th frame to  $\mathbf{b}_{x,y}^i$  at the  $i$  th frame.  $\mathbf{m}_j$  is a motion vector given by  $(m_{j,x}, m_{j,y})$ , and the number of candidates for the motion vectors is

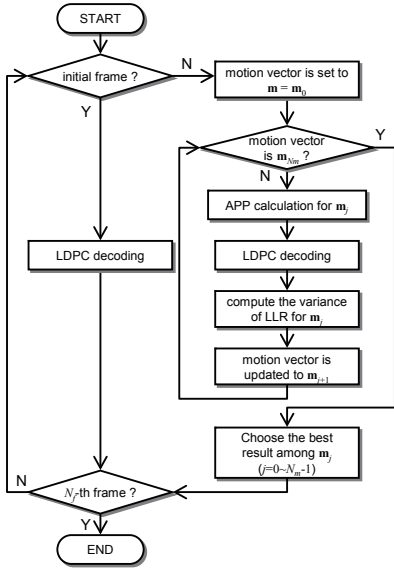


Figure 5: A flowchart of the video frame decoding to recover the transmitted image with side information.

$N_m$ , i.e.,  $j \in \{0, \dots, N_m - 1\}$ . The transition probability  $\Pr(\mathbf{b}_{x,y}^i | \mathbf{b}_{x,y}^{i-1})$  is shown in Fig. 4 where provides probability density on its difference between  $\mathbf{b}_{x,y}^i$  and  $\mathbf{b}_{x,y}^{i-1}$ . This density function is obtained from correlation characteristics between two successive frames among 10 endoscopy video images. The figure shows that an image has high correlation with the image at the previous frame since its cumulative probability over the differences on its pixel value from  $-4$  to  $+4$  reaches 90%.  $\Pr(\mathbf{b}_{x-m_j,y-m_j}^{i-1})$  is the LLR for the previous frame.

A flowchart of these decoding procedure including estimation of the motion vector is shown in Fig. 5. The LLR after the decoding is monitored among the different motion vectors  $\mathbf{m}_j$ . Finally, we decide the maximum likelihood of one denoted as  $\mathbf{m}_{ML}$ ; and then, the recovered bits for the  $i$  th frame image are obtained as

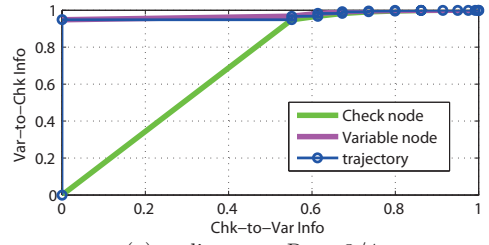
$$\hat{b}_{x,y,k}^i = \begin{cases} 1 & \text{if } L(b_{x,y,k}^i | \mathbf{m}_{ML}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

RGB image data is constructed through demosaicing the recovered data  $\hat{b}_{x,y,k}^i$ .

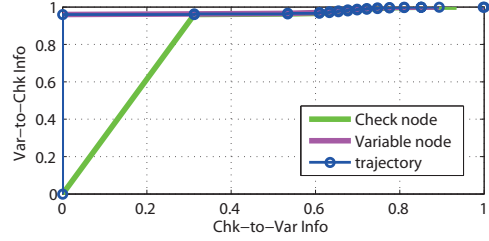
### 3. DESIGN OF THE LDPC CODE AND PERFORMANCE EVALUATION

#### 3.1 Design of LDPC code for the video coding

Its design of the ECC applied into the video encoding is crucial to decide its performance of video compression. In this paper, the use of LDPC is considered because an LDPC encoder is simply implemented through EXOR operations. In addition, LDPC enables provision of only parity bits transmitted in the successive frames; thus the use of LDPC enables omission of the decimation process from the encoder. In order to find a suitable LDPC code, EXIT chart analysis [9] is introduced which is applied into analy-



(a) coding rate  $R_c = 3/4$



(b) coding rate  $R_c = 5/6$

Figure 6: EXIT charts for LDPC codes with coding rate  $R_c$  of 3/4 or 5/6 under an AWGN channel with SNR of 3dB.

sis of iterative decoding in LDPC codes. The EXIT chart is shown in Fig. 6 for LDPC codes with different coding rates  $R_c$ . The length of code  $L_{ecc}$  is set to 64800 bits, and coding rate  $R_c$  is set to either 3/4 or 5/6. These codes are regular LDPC codes, which are adopted in digital video broadcasting (DVB) standard [10]. The convergence behavior obtained from the EXIT charts says that the both LDPC codes achieves error-free conditions, i.e., their mutual information reaches 1. The number of iterations to reach there is 8 and 15 for coding rate  $R_c$  of 3/4 and 5/6, respectively. Since the higher coding rate is capable of realizing better compression rate  $R$  as shown in Fig. 1, the 5/6-rate LDPC code is chosen as ECC in the video encoding. Histograms of its LLR on its decoding iteration under signal-to-noise ratio (SNR) of 1 dB or 3 dB are shown in Fig. 7. In the histograms, a modification is applied that negative LLRs implies correct ones. The results show that iterative decoding for the LLR in SNR of 3 dB is capable of moving its distribution; the encoding and decoding using  $R_c$  of 5/6 is possible to recover the compressed image data transmitted through AWGN channel with SNR of 3 dB. On the other hand, the histogram of the LLR in SNR of 1 dB does not show any shifts of the distribution; is expected that the recovery of its compressed data is not achievable.

#### 3.2 Rate-distortion characteristics

In order to evaluate the performance of the encoding and decoding methods, we carried out simulation using an endoscopy video image. This video image consists of 128 frames taken in a digestive tract. The first- and the last-frame images are displayed in Fig. 8. These images are quantized in 8 bits ( $N_B = 8$ ). Entropy conditioned by the previous frame image is 1.97 bpp, 2.39 bpp, and 2.79 bpp for R, G, and B, respectively. Then, the theoretical lower bound on the compression rate is 1.97 bpp / 8 bit = 24.6 % for R, 2.39 bpp / 8 bit = 29.9 % for G, and 2.79 bpp / 8 bit = 34.9 % for B in the 8-bit quantization case. The average bound among the three color components becomes 29.8%, which is

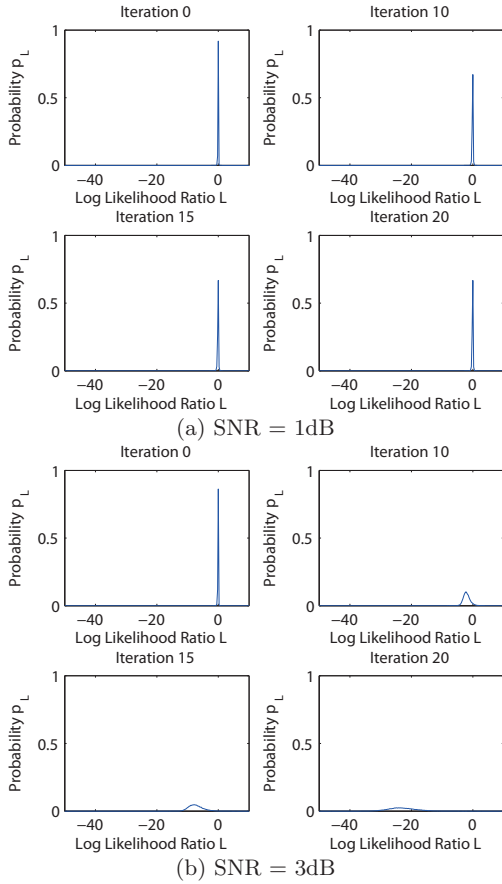


Figure 7: Histograms of LLR at the decoding of the LDPC codes with coding rate  $R_c$  of 5/6 under an AWGN channel with SNR of 1 or 3dB.

the lower-bound on the compression ratio in the proposed method because the method processes the color components together as shown in 2. The system specifications for the simulation are listed in Table 1. The regular LDPC code with  $L_{ecc}$  of 64800 as shown in the previous section, which is adopted in digital video broadcasting (DVB) standard, is used as the channel code. In this simulation, the coding rate  $R_c$  is fixed to 5/6 as shown in the previous section, and the bit rate is controlled by the number of padding bits at the LDPC encoder. We set that both the parity and information bits are transmitted in the first frame; then, for the successive 127 frames ( $N_f = 128$ ), only their parity bits are transmitted. The channel SNR in AWGN ranges from 1 dB or 3 dB, which is viewed as a severe case in capsule endoscope applications [8].

Fig. 9 displays rate-distortion characteristics corresponding to the R, G, and B components in the SNR. The compression rate is defined as the ratio between its required bits per pixel and original bits (= 8); and the distortion is evaluated by the worst root-mean-squared error (RMSE) among the 128 frames. As shown in the figure, non-loss results are achieved in the channel SNR of 3 dB for every color component; but, a residual RMSE of around 15 is observed in the channel SNR of 2 dB. In the channel SNR of 1 dB, any

Parameters	Values
LDPC code	$L_{ecc} = 64800, L_{info} = 54000$ $R_c = 5/6 = 0.833$
Source video image data	Raw (Bayer) data ( $N_B = 8$ ) $x_{max} = 256, y_{max} = 256$ $N_f = 128$
Channel	AWGN
Modulation method	$\pi/2$ -shift BPSK
Search region of $\mathbf{m}$	$3 \times 3$ pixels ( $N_m = 9$ )
$\Pr(\mathbf{b}_{x,y}^i   \mathbf{b}_{x,y}^{i-1})$	The curve shown in Fig. 3
LDPC decoding	Sum-product algorithm

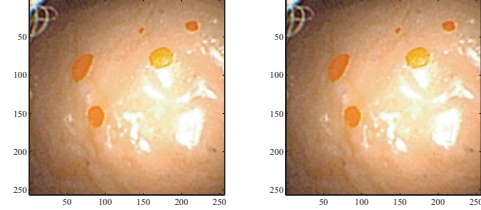


Figure 8: The frame images used in the test for performance evaluation; the left one is the first image, and the right one is the 128th frame. The images consist of 128-frame 8-bit quantized RGB image.

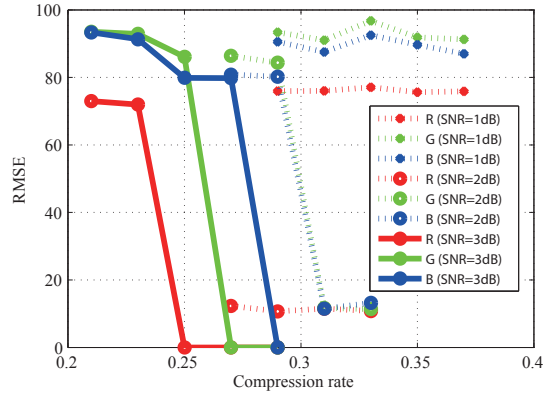


Figure 9: Rate-distortion characteristics of the video coding based on LDPC code. The coding rate is set to 5/6. The SNR in AWGN channel is ranging from 1 to 3 dB.

coding gain is observed as the increase of the compression rate  $R$ . In the SNR of 3 dB, the B component provides a non-loss result when the compression rate  $R$  is above the theoretical bound given by 0.35. This bound is for when encoding and decoding is closed in every color component; however, the decoding procedure presented in this paper estimates the motion vector by using the R component only since the conditional entropy between successive frames is the lowest. This is why such additional gain is as observed in Fig. 9 for the B component. Fig. 10 shows the RMSE corresponding to each frame image in the channel SNR of 3 dB. The results show that the RMSE is accumulated as the number of passed frame images increases, because the

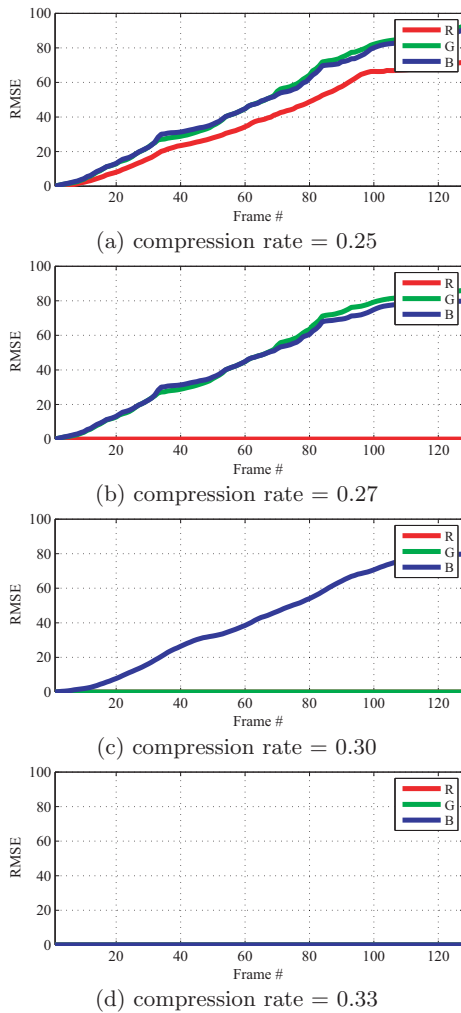


Figure 10: RMSE corresponding to each frame image based on proposed coding with coding rate of 5/6 under SNR of 3 dB.

decoding scheme is based on the frame correlation by using the previous frame. One possible way to reduce the accumulated error is to decrease the number of successive frames  $N_f$ ; though, the decrease of the number of the successive frames  $N_f$  reduces its compression rate  $R$ . For the compression rate of 0.33, BPSK is requested to maintain throughput of 1.84 Mbps on order to provide 10 frames per second.

### 3.3 Comparison with loss-less and lossy JPEGs

JPEG is the most popular compression standard for image data. Here, we make a comparison between the presented method and JPEG. For the JPEG systems, we apply a regular LDPC code with coding rate of 0.5 as their channel code to combat errors caused by the AWGN channel. Their rate-distortion curves are shown in Fig. 11. For the lossy JPEG, the quality index at the compression is set to provide the best quality since its target application is medical diagnoses. As shown in the results, the video coding and

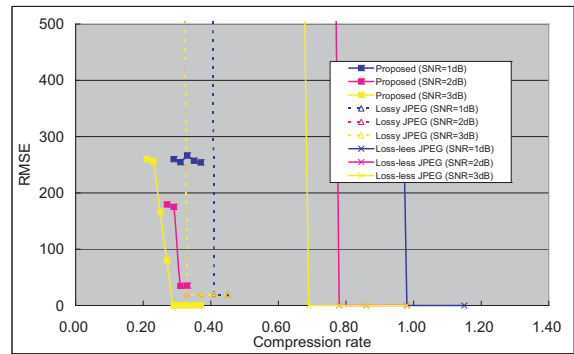


Figure 11: Performance comparison with LDPC-coded JPEGs in term of the rate-distortion characteristics.

decoding provided in this paper give the best performance in the channel SNR of 3 dB. The lossless JPEG requires a compression rate of 0.69 to provide a non-loss result, which is twice that of the proposed coding case. For the lossy JPEG cases, residual RMSE of around 20 is observed. In the channel SNR of 1 and 2 dB, superior RMSE curves are obtained by using an LDPC-coded lossy JPEG, though its complexity in its encoding process is higher than the proposed one since lossy JPEG requires DCT, entropy coding and other computations.

## 4. CONCLUSIONS

This paper presented a compression scheme for video images applicable into capsule endoscopy. The compression scheme, which is based on distributed video coding, is realized by employing a regular LDPC code. The evaluation results using endoscope images show that the compression method provides error-free results under AWGN with channel SNR of 3 dB in compression rate of around 30%. We also demonstrated its advantage through a comparison with JPEG.

## 5. REFERENCES

- [1] B. S. Lewis and P. Swain, "Capsule endoscopy in the evaluation of patients with suspected small intestinal bleeding: Results of a pilot study, Gastrointestinal Endoscopy," vol. 56, no. 3, pp. 349–353, Sept. 2002.
- [2] C. Cheng, Z. Lin, C. Hu, and M. Q.-H. Meng, "A novel wireless capsule endoscope with JPEG compression engine," Proc. of IEEE ICAL 2010, Aug. 2010.
- [3] G. Pan, G. Yan, X. Qiu and X. Song, "A Novel JPEG-based Wireless Capsule Endoscope," Biomedical Instrumentation & Technology, vol.44, no.6, pp.519–522, Nov. 2010.
- [4] Anna N. Kim, Tor A. Ramstad and Ilanko Balasingham, "Very Low Complexity Low Rate Image Coding for the Wireless Endoscope," Proc. of the IEEE 4th Int Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), Spain, Oct. 2011
- [5] Anna N. Kim, Eirik Jenssen Daling, Tor A. Ramstad, and Ilanko Balasingham, "Error Concealment and Post Processing for the Capsule Endoscope," Proc. of

- 7th IEEE/ACM Body Area Networks Conference (Bodynets), Oslo, Norway, Sept. 2012
- [6] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding for motion video," Asilomar conf. on signals, systems and computers, 2002.
  - [7] J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Information Theory, vol. IT-19, pp. 471-480, July 1973.
  - [8] K. Takizawa, T. Aoyagi, K. Hamaguchi, and R. Kohno, "Performance evaluation of wireless communications through capsule endoscope," Proc. of IEEE EMBC'09, Sept. 2009.
  - [9] S. ten Brink, "Convergence of Iterative Decoding," Electronics Letters, 35(10), May 1999.
  - [10] ETSI EN 302 307, "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications," 2005.