31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Low-complexity Video Compression for Capsule Endoscope Based on Compressed Sensing Theory

Jing Wu, Ye Li

Abstract— **Recently, the notions of Compressed Sensing (CS) have attracted attention as an innovative concept in signal processing. In this exploratory paper, a CS-based video compression approach suitable for wireless capsule endoscopy is proposed. In general, the amount of video data generated by capsule endoscopy is so large that video compression is the best way to lower the communication bandwidth and save the RF transmitting power. However, due to power limitation and small size conditions, traditional video compression techniques are not appropriate. Applying state-of-the-art CS theory may significantly reduce power consumption and memory of video compressor, thanks to its low computational complexity. The proposed approach is based on YUV color space conversion, blocking, zigzag scan and CS measuring. Experimental results show the feasibility of the proposed idea and that future improving works are necessary.**

I. INTRODUCTION

Gastrointestinal (GI) endoscope has been widely applied for diagnose of diseases: celiac disease, vascular disorders, Crohn's Disease, benign and malignant tumors of the small intestine, and medication related small bowel injury. Nowadays, there are two classes of GI endoscope: the wired endoscope and the wireless capsule endoscope. The former enables efficient diagnosis using real video and biopsy samples. However, it causes pain and discomfort to patients, making it difficult for the endoscope to push in. To overcome the suffering of patients, the invention of wireless capsule endoscopy (CE) [1] becomes a breakthrough. The CE moves through the internal GI tract with the aid of peristalsis and transmits real video of the intestine wirelessly. This new biotelemetry technique can provide more valuable diagnostic information than conventional one.

However, the amount of data associated with capsule endoscopy video is so large that it may cause significant power consumption in RF transmitter. In applications of capsule endoscope, it is imperative to consider battery life and performance trade-offs. The huge amount of data can be reduced effectively by an image/video compression method that is necessary for saving the power dissipation and bandwidth of RF transmitter.

Traditional image/video compression techniques, including JPEG and MPEG, use inter/intra-frame prediction, DCT, run-length coding, Huffman coding and buffering techniques to earn a good compression ratio to significantly reduce the image bit rate. Some compression algorithms for capsule endoscope based on JPEG and MPEG have been reported to get higher compression ratio [2-5]. But their complexities of the calculation require intensive computation, lots of memory and consume much power from battery. Considering the wireless capsule endoscope, it is not worthwhile to waste power on calculation complexity as long as the compression ratio and quality are acceptable.

Over the past few years, a new framework called as Compressed Sensing (CS), which is also called Compressive Sampling by some other authors, has been developed for sampling and compression. It builds upon the groundbreaking work by Candes et al. [6] and Donoho [7], who showed that by employing linear programming or other mathematical programming methods, a spatially sparse signal can be precisely recovered from only a small set of measurements. The CS principle provides the potential of dramatic reduction of sampling rates, power consumption and computational complexity [8][9]. Due to its great practical potentials, it has stirred great excitements both in academia and industries in the past few years. An available source list of related works can be found at [10]. CS is just suitable for low-power imaging devices (e.g., sensor networks and wireless capsule endoscopy) due to its much lower implementation cost.

Since a large quantity of video frames fit the criterion of CS, we started focusing on applying CS compression techniques for CE video to trade-off battery life and performances.

II. COMPRESSED SENSING THEORY OVERVIEW

The Shannon/Nyquist sampling theorem specifies that to avoid losing information while sampling a signal, at least twice the signal bandwidth should be covered. However, according to CS, only $M(M << N)$ non-adaptive linear measurements of a K-sparse signal of N samples contain sufficient information for perfect reconstruction using non-linear optimization methods, provided that some conditions are satisfied [6][7]. Formally, let $\boldsymbol{x} \in \Re^{\mathrm{N}}$ be a real-valued, finite-length, one-dimensional, discrete-time signal. Any signal in $\Re^{\mathrm{N}}$ can be represented in terms of an orthonormal basis of $\mathrm{N} \times 1$ vectors $\{\boldsymbol{\Psi}_i\}_{i=1}^{N}$. Using

Manuscript received April 23, 2009. This work was supported in part by the Nature Science Foundation of Province Guangdong, China (No. 8478922035-X001220), and supported in part by the Science and Technology Plan Project of Shenzhen, China.

Jing Wu is with Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen 518067, China, and is with Key Laboratory for Biomedical informatics and Health Engineering, Chinese Academy of Sciences, Shenzhen 518067, China (e-mail: wooking@163.net).

the $\mathrm{N}\times\mathrm{N}$ basis matrix $\boldsymbol{\Psi}:=[\boldsymbol{\Psi}_1\,|\,\boldsymbol{\Psi}_2\,|\cdots|\,\boldsymbol{\Psi}_N]$, a signal $\boldsymbol{x}$ can be expressed as

$$\boldsymbol{x}=\sum_{i=1}^{N}\theta_i\boldsymbol{\Psi}_i=\boldsymbol{\Psi\theta} \qquad (1)$$

where θ_i is weighting coefficients. Clearly, if $\boldsymbol{\Psi}$ is full ranked, $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are equivalent representations of the signal: $\boldsymbol{x}$ in the time or space domain, and $\boldsymbol{\theta}$ in the $\boldsymbol{\Psi}$ domain. The signal $\boldsymbol{x}$ has a sparse representation if it is a linear combination of only K basis vectors. That is, only K coefficients of $\{\theta_i\}, i=1,2,\cdots,N$ in (1) are nonzero and the rest N - K ones are zero.

Suppose that we take $M(M<<N)$ linear, non-adaptive measurement of $\boldsymbol{x}$ through the following linear transformation:

$$\boldsymbol{y}=\boldsymbol{\Phi x}=\boldsymbol{\Phi\Psi\theta}=\boldsymbol{\Omega\theta} \qquad (2)$$

where $\boldsymbol{\Phi}$ is an $\mathrm{M}\times\mathrm{N}$ matrix, $M(M<<N)$. $\boldsymbol{x}$ is thus downsampled to a $\mathrm{M}\times 1$ vector $\boldsymbol{y}$. Because $\mathrm{M}<\mathrm{N}$, the task of reconstructing $\boldsymbol{x}$ from $\boldsymbol{y}$ seems ill-conditioned. However, the additional assumption of the sparsity of $\boldsymbol{x}$ makes it possible and practical. The signal $\boldsymbol{x}$ may be exactly reconstructed under the minimum 1-norm reconstruction with high probability, i.e.:

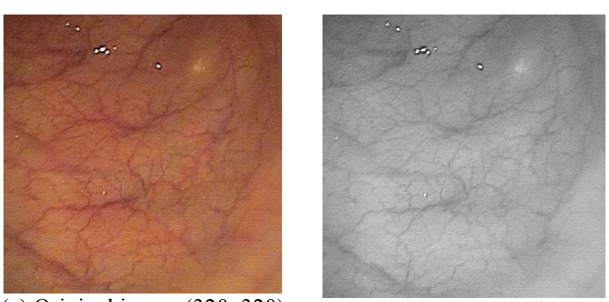
$$\begin{cases}\hat{\boldsymbol{\theta}}=\arg\min\|\boldsymbol{\theta}\|_1\\ \text{s.t. } \boldsymbol{y}=\boldsymbol{\Phi\Psi\theta}\end{cases} \qquad (3)$$

This is a convex optimization problem that can be conveniently reduced to a linear program.

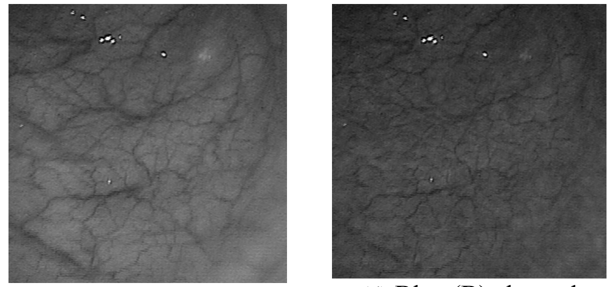
III. CS-BASED VIDEO COMPRESSION FOR CE

A. Color space conversion

CE normally uses CMOS sensor for imaging. CMOS sensor produces four color channels per pixel: one red (R), two green (G) and one blue (B) [11]. Because there is highly redundancy in the information contained in the three channels, as is shown in Fig.1, the color space conversion is the first step in compression.



(a) Original image (320x320) (b) Red (R) channel



(c) Green (G) channel (d) Blue (B) channel

Fig.1 Color image decomposed in its Red, Blue and Green channels: the clear similarity contained in the R, G and B pictures shows huge redundancy present.

Normally YUV color space conversion is done by the following popular transformation:

$$\begin{pmatrix}Y\\U\\V\end{pmatrix}=\begin{pmatrix}0.299 & 0.587 & 0.114\\ -0.147 & -0.289 & 0.436\\ 0.615 & -0.515 & -0.100\end{pmatrix}\begin{pmatrix}R\\G\\B\end{pmatrix} \qquad (4)$$

where Y is luma component while U and V denote blue and red chroma components, respectively. To make the system suitable for a binary treatment, a simplified solution is used to remove floating point calculations. The matrix above is reduced to:

$$\begin{pmatrix}0.250 & 0.500 & 0.125\\ -0.125 & -0.500 & 0.500\\ 0.500 & -0.500 & -0.125\end{pmatrix} \qquad (5)$$

Thus, the simplified transformation only contains additions or subtractions and division by 2 that can be implemented by simple bit shifting.

The evident advantage of applying this color space conversion is illustrated by from Fig.1 to Fig.4. One can easily notice that red, blue and green are distributed from fully white to fully black, showing that one needs the total range of values to code them. But on the other hand, in the U and V channel, the values are all contained in a very narrow range, which results in a very narrow histogram. Due to the higher sparsity, CS can more easily reconstruct them.

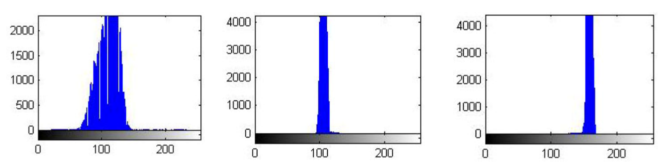
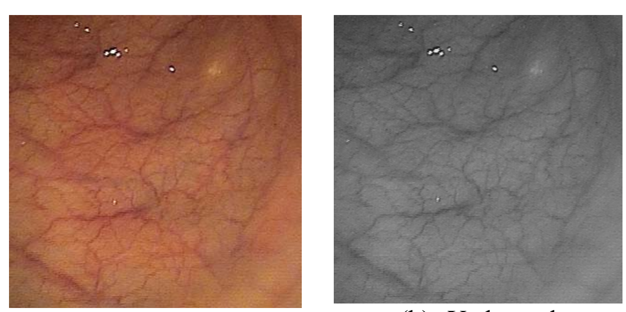
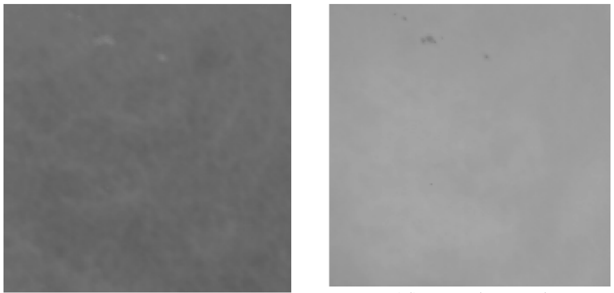


Fig.2. Histograms of the three channels: red (left), blue (mid) and green (right).



(a) Original image (b) Y channel



(c) U channel (d) V channel

Fig.3 Color image decomposed in its Y, U and V channels.

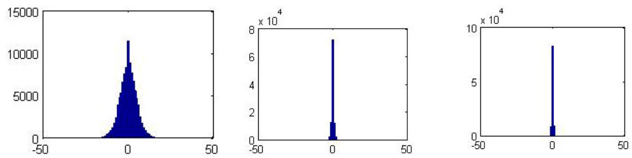


Fig.4 Histogram of the delta between two neighbor pixels in Y (left), U (mid) and V (right) channels.

B. *Blocking size*

The CS reconstruction performance increases as the size grows. However, the computational complexity of this problem is $O(n^3)$. Thus, in practical implementation of CE video decompression, there exists a trade-off between complexity and performance. Dividing image into blocks is necessary, and the block size becomes an important parameter. Small blocking size requires less memory in storage and faster implementation, while large blocking size offers better reconstruction performance. In our practical implementation, the 8x8 block size is used for this compromise.

C. *Zigzag scan*

The sparsity of signal is a necessary condition for CS theory. If one block is scanned row by row, the first pixel of the current row will be adjoint to the last pixel of the previous row. It generates high frequency and consequently cutdown the sparsity of signal. Thus, it is important to select a proper scan model. In this paper, pixels of one block are scanned in the zigzag fashion like in the JPEG standard [12] similar to Fig.5.

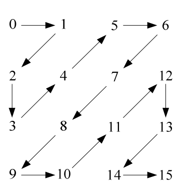


Fig.5 Zigzag scan fashion.

Overall, the function of the video compression is shown in Fig. 6. It starts in the format of Bayer patterns and separately processes YUV channels of the raw input. It blocks and zigzag scans input pixels by 8x8, and then computes CS measurements for transmission. Finally, it reconstructs them.

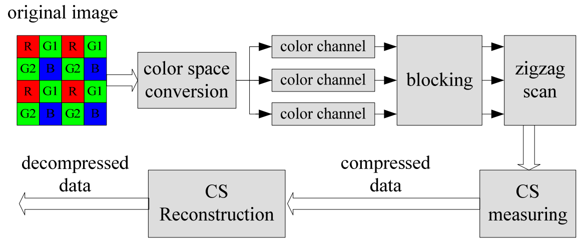


Fig.6 Data flow of the proposed approach.

IV. Experimental Results

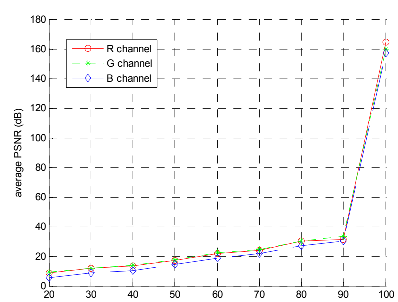
We designed experiments to simulate the compression and reconstruction process on MATLAB platform, with 10 GI color images of 320x320 resolutions. The measurement matrices used were random sequences of real number in the range [-1, 1] with Gaussian distribution. We used the OMP (Orthogonal Matching Pursuit) minimization algorithm, being a common algorithm, in reconstruction and performed. All our experiments used FFT basis.

The sampling percent (SP) is defined as the ratio of the compressed data size to the raw data size (i.e. SP=M/N). The measure of compression quality is the peak signal-to-noise ratio (PSNR), given in equation (6).

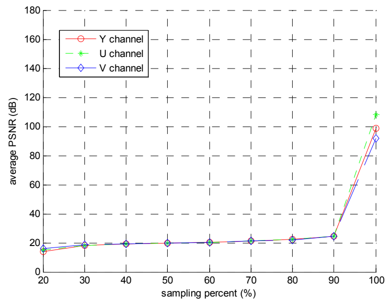
$$PSNR = 10\log_{10}\left(\frac{W \times H \times 255^2}{\sum_{w=1}^{W}\sum_{h=1}^{H}(x_{w,h}^{reconstructed} - x_{w,h}^{original})^2}\right) \quad (6)$$

where W and H denote the width and height of image, respectively. In this paper, W and H are both 320.

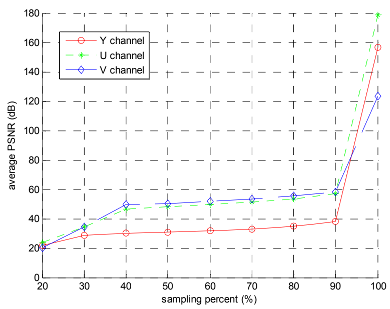
Another color space RGB and blocking size 4x4 are also tested. Average PSNRs of images reconstructed using 8x8-blocking RGB, 4x4-blocking YUV and 8x8-blocking YUV are shown in Fig.7. We note that reconstruction performance of YUV is much better than that of RGB in all SPs from 20% to 100%, especially that of U and V channels. The 8x8-blocking significantly outperformed 4x4-blocking for most of the times. From Fig.7(b) we can further observe that application of YUV instead of RGB does not results in higher performance of U and V channels when using 4x4-blocking, but does significantly higher performance when using 8x8-blocking. Fig.8 shows some reconstructed images under several SPs.



(a) PSNRs of reconstruction of 8x8-blocking RGB

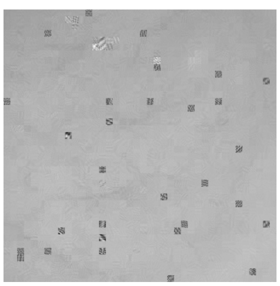


(b) PSNRs of reconstruction of 4x4-blocking YUV

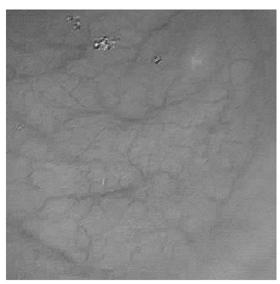


(c) PSNRs of reconstruction of 8x8-blocking YUV

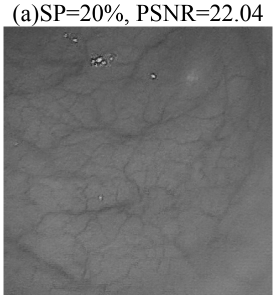
Fig.7 Comparison between the reconstruction performances.



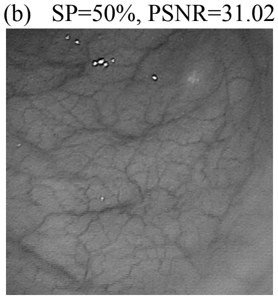
(a)SP=20%, PSNR=22.04



(b) SP=50%, PSNR=31.02



(c)SP=80%, PSNR=34.80



(d) SP=100%, PSNR=156.59

Fig.8 Reconstructed Y channels with various SPs.

V. Conclusions and Future Works

In this paper, we have presented a novel CS-based approach of video compression for wireless capsule endoscopy application. Thanks to CS theory, from the viewpoint of the compression simplicity, it is doubtless that our approach compares favorably with existing more sophisticated compression algorithms including JPEG and MPEG. Our approach is very helpful for the CE processor to obtain less memory and low power. However, the reconstructed image will be severely distorted when sampling percent is fewer than 20%. Hence, further development is necessary for improving reconstruction performance.

As this paper is exploratory, there exist many intriguing questions that future works should consider. First, the theory of block-CS requires to be developed. Secondly, the optimization reconstruction criteria needs to be investigated. Thirdly, it is interesting to exploit the characteristic of gastrointestinal video, for example, the joint sparsity model in simultaneously recovering the R-G-B channels from the compressed measurements.

References

[1] F. Gong, P. Swain, and T. Mills, "Wireless endoscopy," Gastrointestinal Endoscopy, vol.51, no.6, pp.725-729, 2000.

[2] P. Turcza, and M. Duplaga, "Low-power image compression for wireless capsule endoscopy," Proc. of the IEEE Int. Workshop on Imaging Sys. and Tech., pp. 1-4, 2007.

[3] M. Lin, L. Dung, and P. Weng, "An ultra low power image compressor for capsule endoscope," BioMedical Engineering Online 2006, vol. 5:14.

[4] Xiang Xie, GuoLin Li, XiaoWen Li, "A new approach for near-lossless and lossless image compression with Bayer color filter arrays," Third International Conference on Image and Graphics 2004, Hong Kong, pp: 357-360.

[5] Lan-Rong Dung, Yin-Yi Wu, Hsin-Cheng Lai, and Ping-Kuo Weng, "A modified H.264 intra-frame video encoder for capsule endoscope," Biomedical Circuits and Systems Conference, 2008. BioCAS 2008. IEEE 20-22 Nov. 2008 pp: 61-64.

[6] E J. Candes and M B. Wakin, "An introduction to compressive sampling," IEEE Signal Processing Magazine, March 2008.

[7] D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory, vol. 52, pp. 1289–1306, July 2006.

[8] M.B. Wakin, J.N. Laska, M.F. Duarte, D. Baron, S. Sarvotham. D. Takhar K.F Kelly. R.G Baraniuk, "An architecture for compressive imaging," IEEE Image. Proc, pp: 1273-1276, Oct 2006.

[9] Yifu Zhang, Shunliang Mei, Quqing Chen, and Zhibo Chen, "A novel image/video coding method based on compressed sensing theory," Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on March 31 2008-April 4 2008 pp: 1361-1364.

[10] http://www.dsp.ece.rice.edu/cs/

[11] B.E. Bayer, "Color Imaging Array," U.S. Patent 3,971,065, 1976.

[12] G. K. Wallace, "The JPEG still picture compression standard,zz" Com. of the ACM, vol. 34, pp. 30-44, April 1999.