

Specialized Video and Physiological Data Coding System for Remote Monitoring

Qiang Liu¹, Datong Chen⁴, Mingui Sun^{1,2,3} and Robert J. Scwabassi^{1,2,3}

Laboratory for Computational Neuroscience

Departments of ¹Neurological Surgery, ²Electrical Engineering and ³Biomedical Engineering
University of Pittsburgh, Pittsburgh, PA 15261, USA

⁴Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract—Patient monitoring via video and physiological data recording can now be performed outside hospitals. This procedure, usually performed in a prolonged manner, generates a considerable amount of data, which calls for efficient ways for archiving and transmission. In this work, we present a specialized system to code the video and the physiological data recorded from a patient, aiming at a reduced bandwidth requirement compared to the conventional methods. We've developed an object-based approach to coding the monitoring video. By applying two change detection methods, we decompose a video frame into three video object planes (VOPs) representing the background, the stationary foreground and the moving foreground. These VOPs are coded at different frame rates, leading to a reduced overall bit rate. For coding the physiological data (using electroencephalogram, i.e. EEG, as an example), we present an effective solution by using a combination of the lifting scheme and the SPIHT algorithm. This approach is featured with a wavelet-quantization algorithm that enables a scalable transmission. The feasibility of this proposed system is demonstrated by our experimental results.

1. INTRODUCTION

Physiological data recording has been utilized for decades in hospitals for the purpose of patient monitoring. Some monitoring processes (e.g. epileptic patient monitoring) are accompanied by video recording to provide a cross reference for diagnostic evaluations. Recently, thanks to the development of telemedicine technology, these processes can be performed outside hospitals for home care applications. Usually carried out in an extensive amount of time (hours and days), these processes produce a large-scaled data set. There exists a demand for a coding system that is capable of highly efficient transmission and archiving of the monitoring data (including both the video and the physiological data). The existing systems, such as Bio-Logic and the Grass-Telefactor video/EEG systems are mostly based on general-purposed coding standards. These systems are not optimized for the need of coding the prolonged recordings and may not suit the home care applications, where the storage and transmission bandwidth are both very limited. For these reasons, we

develop a video and physiological coding system specialized for the patient monitoring purposes.

In video monitoring, the features of the video should be specifically considered in the coding scheme. These features include, 1) hardly any global motion present in the video, 2) recording environment usually remaining unchanged, and 3) for most of the time, only small movements are present, such as those caused by body parts. Higher coding efficiency is expected if these features are utilized specifically in the design of a compression engine. For instance, the background regions should be coded with much relaxed quality (both spatial and temporal), in contrast to the foreground (i.e. the patient), especially the moving part of the foreground. In the light of MPEG-4, these considerations can be formed into an object-based coding scheme, which decomposes a frame of a monitoring video into three video object planes (VOPs), one representing the recording environment (denoted by VOP1), one representing the stationary foreground (VOP2), and the other the moving foreground (VOP3). These VOPs are coded at different frame rates, with VOP3 the highest and VOP1 the lowest. Since VOP3 is usually small in size (in pixels), the overall bit rate can be reduced, while an essential quality on the foreground is preserved.

For the coding of the physiological data, we adopt a device-adaptive strategy that allows a scalable transmission. This strategy was motivated by a fact that a considerable redundancy with respect to the data resolution may exist in the transmission. In most diagnostic evaluations, the physiological waveforms are visually examined. The display resolution of these waveforms, however, is often lower than that of the data. This problem is particularly significant when portable devices, such as PDA or cell phone, are utilized to display the waveforms. Therefore, in many cases, a scalable transmission that suits the terminal device is desirable. We combine the SPIHT codec, the lifting scheme and wavelet quantization to tackle this problem. In this approach, the physiological data are transformed into wavelet coefficients by lifting scheme. These coefficients are quantized so that the resolution of the reconstructed data matches the remote display terminal.

We present a preliminary system that implements the above coding schemes. The experimental results on

This work was supported in part by National Institutes of Health Grant No. NS/MH38494 and EB002309.

real-world data suggested a considerable improvement over the conventional coding systems for remote monitoring.

2. OBJECT-BASED CODING FOR MONITORING VIDEO

In our strategy, three video objects are defined based on the features of the monitoring video. A snapshot of the recording environment (i.e. the monitoring room) forms the VOP1. If the camera is allowed to pan and tilt, the background scene can be updated online [6]. Since the background rarely changes, the VOP1 is coded at very small frame rate, e.g. once per minute. The second video object, VOP2, is obtained by constituting the distinction between VOP1 and a video frame. Essentially, VOP2 is consisted of the patient and the changes of the background brought by the patient. The VOP2 is coded at a properly chosen interval, e.g. every second. Within this interval, VOP2 can be considered as stationary foreground regions. The dynamics within this interval are captured by the VOP3, constructed from multiple consecutive video frames by exploring the motion information. Therefore, the VOP3 is usually consisted of the moving body parts of the patient.

To construct the VOP2, we developed an illumination invariant change detection method [1] by combining a simple illumination model (“shading model”) and a statistical test approach. This approach is robust against device noise and illumination variation and easy to implement. A sample result is shown in Fig. 1. The segmentation results can be improved, e.g. removing the holes, by a simple post-processing step.

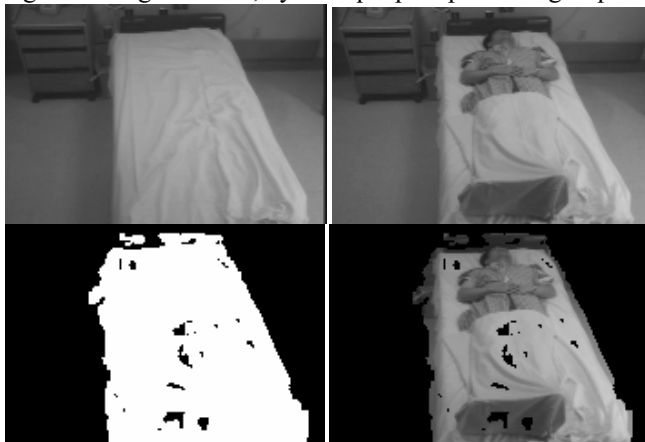


Fig. 1. Top-left: the VOP1, a background snapshot; top-right: a sample video frame; bottom-left: change detection mask, with the white regions denoting the “changed” and the black denoting the “unchanged”; bottom-right: VOP2 that is consisted of the patient and a deformed bed. The holes in the change detection mask can be compensated by morphological operations.

For VOP3, we developed a multiple-frame-based change detection method [2], which identifies moving pixels across a group of video frames. This method is more sensitive to small motions compared to traditional methods using frame pairs. A typical example is shown in Fig. 2, where the results of our method and a classic method (by Aach [3]) are compared.

As these three VOPs have different activity levels, they are generated at different time and thus coded at different frame rates. The VOP1, VOP2 and VOP3 have life-spans of T_1 , T_2 and T_3 respectively, as shown in Fig. 3. The values of T_1 , T_2 and T_3 are determined experimentally and T_3 is just the interval of two frames.

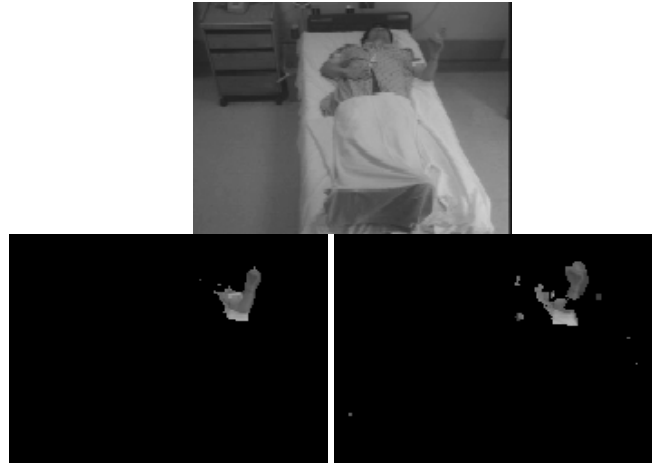


Fig. 2. Top: a video frame that contains a moving hand; bottom-left: VOP3 detected by the multiple-frame-based change detection; bottom-right: change detection result of significance test method [3].

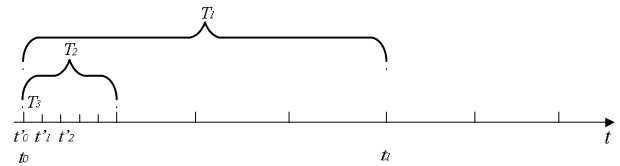


Fig. 3. The life-span of the VOPs. VOP1, VOP2 and VOP3 are coded at the intervals of T_1 , T_2 and T_3 respectively. T_3 is just the interval of two frames.

In the decoding process, the time stamps are extracted and the video frame is reconstructed by means of overlaying: VOP3 is surmounted on top of VOPs and then on VOP1. Since these VOPs are coded at different frame rates, they are rendered at different time points. Accordingly, VOP1 and VOP2 are held for intervals of T_1 , T_2 respectively after being decoded.

3. PHYSIOLOGICAL DATA CODING

We combine the 1D SPIHT [5] and the lifting scheme [4] to code the physiological data. The lifting scheme is a state-of-the-art method to implement the wavelet transform. It utilizes a ladder-like structure to compute wavelet transform, rather than the traditional filter bank. Because of this structure, the lifting scheme is more computationally efficient and can realize invertible integer-to-integer transform. The SPIHT is a well-known codec that is performed in wavelet domain. The wavelet coefficients are coded bit plane by bit plane, namely, the most significant bits of all the coefficients are output to the bit stream first, then the less significant bits. The coding is

lossless if all the bit planes of the entire set of coefficients are coded. However, as mentioned previously, there exists a typical mismatch between the data resolution and the display resolution. In many cases, a lossless transmission is simply a waste of bandwidth. Therefore, we developed a device-adaptive approach to coding the physiological data.

Let us consider the “vertical resolution” of the display window, which is the number of pixels available in the vertical direction. This resolution is often much less than the range of the data value. For example, EEG samples are usually 12-bit integers, ranged from -2048 to 2047, while a laptop screen has normally 768 pixels in the vertical dimension. In addition, physiological data is often recorded in multiple channels, thus each channel is allocated a small portion of the vertical pixels. Obviously, this redundancy can be reduced by transmitting only a portion of the bit planes. The key question is how to determine the quantization step.

Let the data be B -bit integers, then the value range is $A = 2^B$. Let the vertical dimension be n , then a single pixel represents a sub-range of $\frac{A}{n}$. Let b be the number of bit planes (the least significant ones) to be omitted. Let e be the quantization error in the original domain. If b is properly determined, then e produces no visual effect in the display of the reconstructed waveform, which satisfies $e < \frac{A}{n}$. We characterize e in a statistical way: by assuming the distribution of the quantization residuals in the wavelet domain, we calculate the standard deviation of the reconstructed error e . After a multi-level wavelet decomposition, most coefficients are the detail coefficients. We only consider the error generated by quantizing the detail coefficients. Let d_j be the quantization residual of the detail coefficients at level j , and let σ_{d_j} be its standard deviation.

It can be shown that for the lifting implementation of symmetric interpolating filters [4], we may have $\sigma_e \leq \alpha \sigma_{d_j}$,

where σ_e is the standard deviation of e . If we quantize the detail coefficients of J levels,

$$\alpha \approx \sqrt{\frac{J(1 + \sum_k \tilde{g}_e(k)^2)}{2}}$$

where $\tilde{g}_e(k)$ denotes the coefficients of the bandpass wavelet filter \tilde{g} [4]. Assuming d_j is uniformly distributed, we have

$$\sigma_{d_j} = \sqrt{\frac{2^b(2^{(b-1)} - 1)}{3}}$$

One may use $m\sigma_e$ as the measure of the error e , where m is a chosen constant. Then the quantization step 2^b can be numerically obtained by

$$\text{solving } m\alpha \sqrt{\frac{2^b(2^{(b-1)} - 1)}{3}} < \frac{A}{n}.$$

We evaluated this quantization method by testing a large scale EEG data. The quantization errors vs. the number of omitted bit planes (i.e. b) are shown in Fig. 4. Each plot shows the reconstructed error generated from the quantization of the coefficients on a single level. In addition to the standard deviation of the quantization error (in dash lines), the maximum error (in solid line) is also shown in Fig. 4, which is approximately three times the standard deviation. Our measure, i.e. $\alpha\sigma_{d_j}$, is plotted in a thick solid line, serving as an upper bound of the σ_e , which is compatible with our derivation.

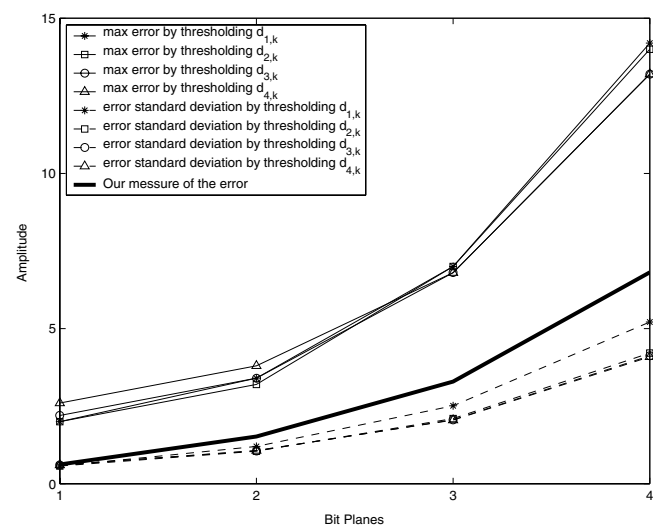


Fig. 4. The standard deviation (dash lines), the maximum (solid lines), and the derived estimate (the thick line) of the error of the reconstructed sequence vs. the number of the omitted bit plane. Each plot shows the error from the quantization of the detail coefficients on a single level. The derived estimate serves as an upper bound of the reconstructed error, which is compatible with the derivation.

4. SYSTEM EVALUATION

In our test system, we utilized Microsoft ConferenceXP and Directshow APIs to implement the video coding. We focused on designing the modules that implement video object construction to be embedded as add-on components in the platform. For simplicity, we used RTTP as the video transmission protocol and FTP to transmit EEG data. In this case, the EEG sequences were coded to files, each containing an epoch defined as one second data of all the channels. We used the time stamps that labeled the video stream as the file names of the EEG files, which enabled the synchronization of the EEG and the video stream.

The experimental video was in the 4SIF format (i.e. 704×480 pixels) recorded by a JVC DV3000 camcorder. Samples of constructed VOP2 and VOP3 and the rate-distortion plots at constant bit rates are shown in Fig. 5.

We compare the coding efficiency of our object-based approach and the conventional frame-based compression by examining the PSNR of the *foreground* regions. It is seen that the object-based method provided a superior quality of the foreground to that by the conventional method. This is because the frame-based method allocated more bandwidth on the background regions, which is a waste in this application.

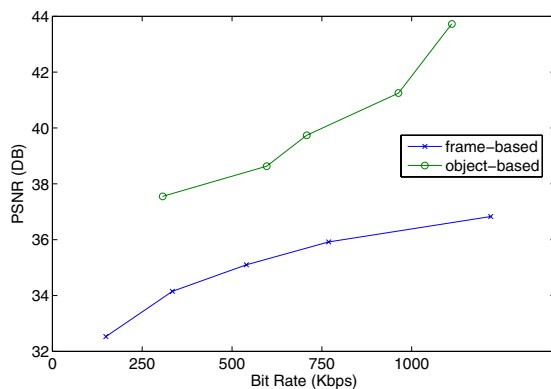
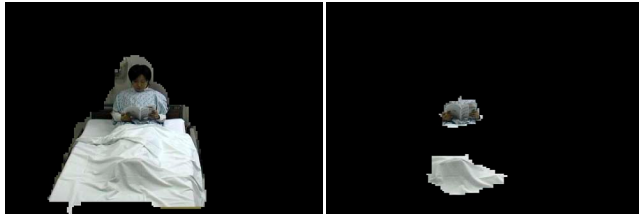


Fig. 5 Experimental results on object-based coding of the patient monitoring video. The top panels show a sample of constructed VOP2 (left) and VOP3 (right) respectively. The rate-distortion plots of both the object-based method (in circle plot) and a conventional frame-based one are compared. We compare the quality (in terms of PSNR) of the foreground regions for both methods. We see that the object-based method provided a superior quality of the foreground to that by the conventional method.

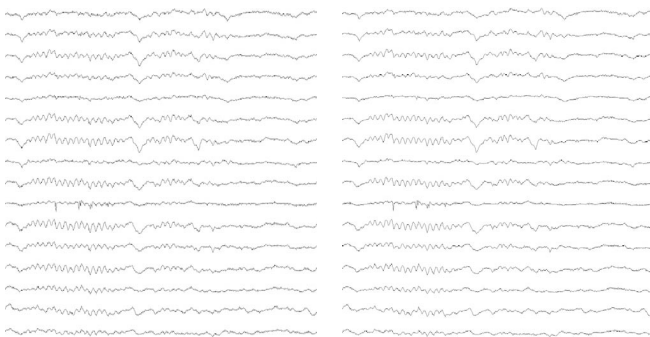


Fig. 6. Screen dumps of two windows displaying 16 channels of EEG segments. The left panel shows a lossless transmission at 38.4 Kbps; the right panel shows a reconstructed data from a lossy transmission at 9.9 Kbps. There are hardly any visible distortions observed from the lossy sequence.

For the physiological data, we tested a set of multi-channel EEG sequences recorded at the University of Pittsburgh Medical Center. Concerning the display qualities, we show in Fig. 6 the screen dumps of the lossless sequence and the

sequence reconstructed from quantized wavelet coefficients. At virtually the same display quality, the bit rates were reduced from 38.4 Kbps to 9.9 Kbps for coding the 16 channels of EEG.

5. CONCLUSION

We've presented a specialized video and physiological data coding system for patient monitoring. An object-based approach has been proposed to advance the video coding efficiency. We showed that a three-object representation of the monitoring video may be utilized to launch content-driven applications. The underlining concept is to selectively code the video content so as to reduce the overall bandwidth requirement, and facilitate future archiving and retrieval functions. Experimental results on a variety of monitoring video sequences showed that the object-based scheme outperformed the frame-based coding in a wide margin. In addition to the video coding, we've also presented a device-adaptive approach to coding physiological waveforms. Considering that the waveforms are usually diagnosed visually on a display device, we propose to transmit the physiological data at the resolution that matches the resolution available at the display terminal. This redundancy reduction may be significant for displaying waveforms on portable devices, such as PDA and cell phones, where both the transmission bandwidth and the display resolution are very limited. A wavelet quantization scheme has been provided along with analytical derivation of the proper quantization step. Experimental results on a large scaled EEG data set validated the proposed scheme. We believe both the video and physiological coding schemes presented in this work would enlighten some health care applications both in- and outside hospitals.

REFERENCES

- [1] Qiang Liu, Sciabassi, R.J.; Mingui Sun, "Illumination-invariant change detection model for patient monitoring video," in *Proc. 26th EMBC*, 2004, vol. 3. pp.1782 – 1785.
- [2] Qiang Liu, Sciabassi, R.J., Mingui Sun, "A new change detection method and its application to epilepsy monitoring video", in *Proc. IEEE 30th Annual Northeast Bioengineering Conference*, 17-18 April. 2004, pp. 59 – 60.
- [3] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," in *Signal Processing*, vol. 31, pp. 165–180, 1993.
- [4] I. Daubechies and W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps," *J. Fourier Anal. Appl.*, Vol. 4, No. 3, pp. 247-269, 1998.
- [5] Amir Said, William A. Pearlman, "A new, fast, and efficient image codec based on SPIHT," *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 6, No. 6, pp. 243-250, June 1996.
- [6] Shao-Yi Chien, Shyh-Yih Ma, Liang-Gee Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 12, No. 7, pp. 577-586, June 2002.