THE COMPARISON OF SSD ALGORITHM WITH OTHER ECG SAMPLING ALGORITHMS

HSIAO-HSUAN CHOU¹, YU-CHIEN SHIAU^{1,2}, TE-SON KUO^{1,3}

¹Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

²Department of Nuclear Medicine, Far Eastern Memorial Hospital, Panchiao, Taipei, Taiwan

³Institute of Biomedical Engineering, National Taiwan University, Taipei, Taiwan

ABSTRACT

We had proposed a novel and fast Electrocardiogram (ECG) signal compression algorithm for non-uniform sampling in time domain [1]. It meets the real-time requirement for clinical application. Moreover, the compression performance is stable and uniform even for abnormal ECG signals. A criterion called sum square difference (SSD) is defined as an error test equation. The algorithm using SSD to calculate error tolerance is applied to the records in MIT-BIH database (with 11-bit resolution and 360 Hz sampling rate). It belongs to the threshold-limited algorithm but [1] does not mention much about this kind of algorithm. In this paper we provide more comparisons among SSD, Fan, scan-along polygonal approximation (SAPA), maximum enclosed area (MEA), and optimization algorithm (OPT) using the two measures called sample compression ratio (SCR) and percent root mean squared difference (PRD) with proper mean offset that [1] does not adopt. The results show SSD outperforms the mentioned algorithms with the same computational complexity $O(n)$. Moreover, the comparison with the best but time-consuming coder OPT ($O(n^3)$) shows how much the algorithm can be improved.

Biomed Eng Appl Basis Comm, 2006(June); 18: 124-127.

Keywords: ECG; sampling; PRD; SCR; SSD

1. INTRODUCTION

To introduce the SSD algorithm, we adopted some descriptions in [1] for readers' convenience and understanding.

There are many algorithms for ECG signal compression to reduce the data size for storage or transmission. They can be classified into at least three major categories [2].

1. Parameter extraction techniques: such as vector quantization (VQ) [3].

2. Transform-domain techniques: such as wavelet transform [4].

3. Direct time-domain techniques: such as SAPA [5], Fan [6], MEA [7], OPT [8], and SSD [1].

This paper focuses on the third category, the time-domain sampling methods. For example, the SAPA and Fan algorithms limit the error within a threshold. The MEA algorithm limits the area error within a specified threshold with less phase delay than Fan, but its PRD is sometimes worse than Fan. The complexity of these algorithms is $O(n)$. There is no description about the relation between the error threshold and PRD in these algorithms - given the PRD, the error threshold is found by trial and error.

The best sample sets could be obtained through the optimization algorithm [8] but it needs huge

Received: Jan. 9, 2006; Accepted: Feb. 27, 2006

Correspondence: Hsiao-Hsuan Chou, Ph.D. Candidate
Department of Electrical Engineering, National Taiwan
University, Taipei, Taiwan

E-mail: hh.hhchou@gmail.com

computation with complexity $O(n^3)$ because it tries all possibilities exhaustively.

2. ADAPTIVE SAMPLING ALGORITHM

2.1. SCR and PRD

In order to compare the performance of different sampling algorithms, two measures are usually adopted: the SCR and PRD. SCR is defined as

$$\text{SCR} = \frac{\text{Number of samples in original data}}{\text{Number of retained samples}}. \quad (1)$$

PRD is defined as

$$\text{PRD} = \sqrt{\frac{\sum_{n=1}^N [y(n) - \hat{y}(n)]^2}{\sum_{n=1}^N [y(n) - \bar{y}]^2}} * 100\%, \quad (2)$$

where $y(n)$ is the original signal of length N with mean $\bar{y}$, and $\hat{y}(n)$ is the reconstructed signal. In [1] PRD is defined without any offset. As a result, it will produce smaller PRD values. For meaningful and fair comparison, it is better to adopt (2) with mean offset.

2.2. SSD

We define SSD to estimate the error of each ECG segment as following

$$\text{SSD} = \sum_{n=n_i}^{n_f} [y(n) - \hat{y}(n)]^2, \quad (3)$$

where $y(n)$ is the original data, $\hat{y}(n)$ is the reconstructed data from linear interpolation between samples, and n_i and n_f are the initial and final data indexes of the segment respectively. As the data index n_f increases, SSD may increase or decrease. The proposed algorithm incrementally grows the segment and increases n_f until we first obtain

$$\text{SSD} > d, \quad (4)$$

where d is an adaptive threshold that will be described in the next section. Once (4) is satisfied, the last data point of the current segment is chosen as a sample, and this sample is treated as the initial point of the next segment. The above sampling procedure is repeated until the end of entire ECG signal.

2.3. Adaptive Threshold d

How to d choose is the key to the proposed algorithm. Consider one ECG signal where the data points are indexed from 1 to N . Let M be the mean square difference:

$$M = \sum_{n=1}^N [y(n) - \bar{y}]^2 / N. \quad (5)$$

Then PRD could be written as

$$\text{PRD} = \sqrt{\frac{\text{SSD} / N}{M}}. \quad (6)$$

Thus d may be estimated roughly by the

$$d > \text{PRD}^2 * M * N. \quad (7)$$

In practice, M and N could be estimated from the first three or more seconds of the ECG signal.

3. RESULT AND DISCUSSION

The algorithms were applied to MIT/BIH database to check the performances.

Table I shows some of the experimental results of SAPA, Fan, MEA, SSD, and OPT algorithms. The PRD performances of SSD are better than the other three threshold-limited algorithms SAPA, Fan, and MEA. The popular Fan is chosen for detailed comparisons.

By adjusting the max error threshold d of SSD and error range ε of Fan [6], respective SCR and PRD pairs of the two algorithms were obtained. According to the experimental results in Table II, the algorithm using SSD is superior to the Fan algorithm with smaller PRD.

Table I. PRD resulted from SAPA, MEA, Fan, SSD, and OPT algorithms

Record	Data Point	SCR	PRD [5] (SAPA)	PRD [7] (MEA)	PRD [6] (Fan)	PRD [1] (SSD)	PRD [8] (OPT)
100_0	1-5400	20.0	25.5%	18.0%	18.0%	14.2%	9.4%
100_1	1-5400	20.0	42.4%	22.8%	20.1%	15.4%	10.3%
119_0	1-540	20.0	23.9%	10.1%	13.3%	9.1%	5.8%
234_0	1-540	20.0	16.5%	13.2%	14.0%	10.6%	7.4%

Table II . PRD resulted from Fan and SSD algorithms

Record	Data Point	SCR	PRD [6] (Fan)	PRD [1] (SSD)
100_0	1-5400	20	18.0%	14.2%
100_0	1-540	20	17.4%	14.9%
100_1	1-5400	20	20.1%	15.4%
100_1	1-540	20	24.0%	20.1%
118_1	1-5400	18	45.8%	35.8%
118_1	1-540	9	20.7%	20.3%
119_0	1-5400	22.7	13.5%	9.9%
119_0	1-540	20	13.3%	9.1%
200_0	1-5400	20	10.3%	9.4%
200_0	1-540	20	13.8%	12.0%

Record 119 is taken for an example to illustrate more details of the algorithms. Fig. 1 shows the sample points of record 119 obtained by Fan with SCR 20 and PRD 13.3%. Fig. 2 shows the sample points of record 119 obtained by SSD with SCR 20 and PRD 9.1%. Fig.3 shows the sample points of record 119 obtained by OPT with SCR 20 and PRD 5.8%. Fig.4 shows the PRD difference between Fan and SSD with various SCR. It is clear the larger the SCR, the larger the PRD difference between SSD and Fan algorithms.

Fig. 1. Fan applied to 119_0 data points from 1 to 540, SCR=20, PRD=13.3%

Fig. 2. SSD applied to 119_0 data points from 1 to 540, SCR=20, PRD=9.1%

Fig. 3. OPT applied to 119_0 data points from 1 to 540, SCR=20, PRD=5.8%

Fig. 4. PRD for SSD and Fan algorithms applied to MIT-BIH 119_0 data points from 1 to 5400 with various SCR (+: Fan, o: SSD)

4. CONCLUSION

This paper demonstrates the performance of SSD is better than the SAPA, Fan, and MEA algorithm. The PRD of SSD is better (smaller) than that of Fan algorithm especially for large SCR cases. Moreover, samples by SSD preserve more diagnostic information than Fan with the same SCR. Although the optimization algorithm provides better PRD, it could not meet real-time requirement, and its computational complexity is $O(n^3)$, far exceeds that of our algorithm $O(n)$.

Moreover, considering to abnormal ECG signals, those are more important than normal signals for diagnosis, the compression qualities degrade when applying the parameter extraction or transformed-domain techniques (the first or the second category) to them. However, our approach wouldn't suffer from this effect because it doesn't need to extract some patterns, periods, or frequencies.

In addition, if necessary, the SCR could be improved by applying some of the quantization or digital source coding methods to the sampled data.

Finally, our algorithm can be applied to not only ECG but also all one-dimensional signals.

REFERENCE

1. H.H. Chou, Y. C. Shiau, and T.S. Kuo: An adaptive real-time ECG sampling algorithm using sum square difference. IEEE the 12th International Conf. on Biomed. Eng. (ICBME) 2005; 187.
2. J. J. Wei, C. J. Chang, N. K. Chou, and G. J. Jan : ECG data compression using truncated singular value decomposition. IEEE Trans. Inform. Technol. Biomed., 2001; 15.4: 290-299.
3. B. Wang, and G. Yuan (1997): Compression of ECG data by vector quantization. IEEE Eng. Med. Biol. Mag., 1997; 16.4: 23-26.
4. B. A. Rajoub: An efficient coding algorithm for the compression of ECG signals using the wavelet transform. IEEE Trans. Biomed. Eng., 2002; 49: 355-362.
5. M. Ishijima, S. B. Shin, G. H. Hostetter, and Y. J. Sklansky: Scan-along polygonal approximation for data compression of electrocardiograms. IEEE Trans. Biomed. Eng., 1983; 11: 723-729.
6. R. C. Barr: Adaptive sampling of cardiac waveforms. J. Electrocard. 1988; 21: 57-60
7. A.C. D'ambrosio, A. Ortiz-conde, and F. J. Sanchez: Percentage area difference (PAD) as a measure of distortion and its use in maximum enclosed area (MEA), a new ECG signal compression algorithm. 4'th IEEE Int. Caracas Conf. on Devices, Circuits Syst., 2002; 1035-1-1035-5.
8. D. Haugland, J. G. Herber, and J. H. Husoy: Optimization algorithm for ECG data compression. Med. Biol. Eng. Comput., 1997; 35: 420-424.