A Constrained Two-Layer Compression Technique for ECG Waves

Kyungguen Byun[1], Eunwoo Song[1], Hwan Shim[2], Hyungjoon Lim[2], and Hong-Goo Kang[1]

Abstract— **This paper proposes a constrained two-layer compression technique for electrocardiogram (ECG) waves, of which encoded parameters can be directly used for the diagnosis of arrhythmia. In the first layer, a single ECG beat is represented by one of the registered templates in the codebook. Since the required coding parameter in this layer is only the codebook index of the selected template, its compression ratio (CR) is very high. Note that the distribution of registered templates is also related to the characteristics of ECG waves, thus it can be used as a metric to detect various types of arrhythmias. The residual error between the input and the selected template is encoded by a wavelet-based transform coding in the second layer. The number of wavelet coefficients is constrained by pre-defined maximum distortion to be allowed. The MIT-BIH arrhythmia database is used to evaluate the performance of the proposed algorithm. The proposed algorithm shows around 7.18 CR when the reference value of percentage root mean square difference (PRD) is set to ten.**

I. INTRODUCTION

As the demand for continuous monitoring of health conditions has increased, so too has the importance of electrocardiogram (ECG) waves due to their non-invasive ability to detect cardiovascular disorders. However, ECG data stored for hours or days becomes redundant because the shape of the ECG is repeated with P, QRS complex, and T waves. In other words, ECG data can be compressed significantly if the periodic characteristic is exploited. Furthermore, the compression technique is more beneficial if the coding parameters can be used directly for diagnosis.

Conventional ECG compression algorithms are categorized into three types: direct, prediction-based, and transform-based coding. In direct coding schemes, redundant samples are discarded such that the reconstruction error between an input and an interpolated signal does not exceed a pre-defined threshold [1-3]. For medical purposes, these algorithms are not suitable because of large reconstruction errors, especially the region of T and P waves [4]. The prediction-based coding schemes that utilize the periodic property of ECG waves represent input signals using linear prediction (LP) or long-term prediction (LTP) coefficients [5][6]. Since normal ECG waves are periodic, they can be modeled by a small number of coefficients, which results in high compression ratios (CR) with small reconstruction errors. However, performance drops significantly if the input ECG signal has irregular shapes mainly caused by arrhythmias. A template-based coding scheme that replaces the long-term predictor with a representative form of ECG wave [7][8] has advantages from a diagnostic perspective because the extracted templates can be viewed as morphologically distinct features. However, similar to the prediction-based coding scheme, the reconstruction error increases to the arrhythmic input beats. In transform-based coding schemes, input ECG waves are compressed by a portion of the coefficients in the transform domain [9][10]. Although they provide high performance even in the case of irregular wave shapes, they need to be decoded first before performing any diagnostic processes. In other words, the coding parameters are not meaningful for diagnosis because they are selected based on the criterion of minimum mean square error (MMSE) between the original and the reconstructed signals. Note that the MMSE criterion does not consider the regional importance of ECG waves; thus, it is difficult to emphasize any meaningful region for diagnosis.

To alleviate the aforementioned limitations, this paper proposes a constrained two-layer coding algorithm that not only decreases the reconstruction error in medically important regions but also represents diagnostic information using coding parameters. In the first layer, the most similar template to the single ECG beat is selected from the registered codebook. If the similarity between the input and the selected template is low, the input signal is registered into the codebook under the assumption that it is a new type of template. By using the information extracted from the first layer, such as shape and distribution of templates, it is possible to detect arrhythmias without performing a decoding process. In the second layer, a discrete wavelet transform (DWT) is introduced to encode the remaining error signal from the first layer. The number of coefficients to be encoded is determined in such a way that overall distortion does not exceed the pre-defined threshold. The proposed algorithm also introduces a template-masking method that is designed to automatically emphasize medically important regions.

II. CONVENTIONAL TEMPLATE-BASED CODING

A typical template-based coding scheme consists of the following four steps: pre-processing, basic unit selection, template registration, and coding process [7][8]. In [7], after segmenting input ECG waves into basic units such as P, QRS complex, and T waves, the templates for each basic unit are pre-trained by fifty periods of ECG using an MMSE criterion. Although this approach shows good performance when the length of the ECG is relatively short or the input heartbeat is regular, there are several limitations in performing

---

This work was supported by Samsung Electoronics

[1]K. Byun, [1]E. Song, and [1]H. Kang are with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea (e-mail: piemaker90@dsp.yonsei.ac.kr, sewplay@dsp.yonsei.ac.kr, hgkang@yonsei.ac.kr).

[2]H. Shim, and [2]H. Lim are with Digital Media & Communication R&D Center, Samsung Electronics Co., Ltd., Suwon, Korea (e-mail: hwan.shim@samsung.com, hyungjoon48.lim @samsung.com).

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

a template registration process. First, the templates obtained over limited time periods cannot be generalized to long-term monitoring applications because signal characteristics vary over time. Second, it is not adequate to use the MMSE criterion in the template registration process because that criterion does not consider any diagnostic information. The algorithm proposed in [8] uses basic units similar to those used in [7]. In this method, the input ECG is compressed with a small number of chosen parameters not to lose diagnostically meaningful features such as RR and PR intervals. However, because the modeling of each basic unit is independent, discontinuities between the modeled template region and non-modeled regions such as TP segments cannot be avoided.

To overcome the limitations described above, three things should be considered. First, the template codebook must be updated continuously in case the signal characteristics are changed or abnormal ECG waves are detected. Second, it is better to introduce the metric of normalized cross correlation (NCC) to register and to compare the templates because this metric allows the similarity of shapes to be measured regardless of their magnitude variation. Third, the basic unit of template should cover the entire single period of the ECG wave in order to avoid discontinuities in the reconstructed signal.

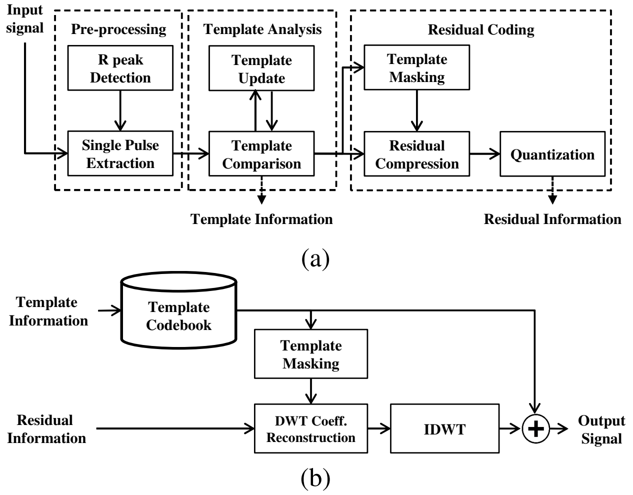
III. PROPOSED ECG CODING METHOD

Fig. 1 depicts the overall framework of the proposed algorithm. In the pre-processing module, a single pulse of the ECG wave is extracted from the input signal, from which noise components are removed before starting the extraction process. The extracted signal is then compared with the registered templates in the codebook to find the closest match. If the distance between the extracted signal and the closest template is large, the extracted signal is registered into the codebook by treating it as a new type of template. The residual signal, obtained by subtracting the selected template from the input signal, is encoded by the DWT to ensure further improvement of coding accuracy. In the decoding process, the residual signal is first reconstructed by performing an inverse DWT (IDWT) to the quantized coefficients decoded from the bitstream. A reconstructed signal is then obtained by adding the residual signal and the matched template which is re-interpolated to original RR interval.

A. Pre-processing module

The pre-processing module first removes noise components such as baseline wandering and power noise. Baseline wandering is a type of noise that is slowly changing DC bias mainly caused by respiration [11]. To remove the noise components, a fifth order of Butterworth band-pass filter with a cut-off frequency of 5-50 Hz is used.

A basic unit of the template is defined by a single period of the ECG. Since an R peak has the maximum amplitude of a single period in general, it is used as a reference point to extract the single period of the ECG wave. Fig. 2 shows an example result of R peak detection algorithm. The searching range of the R peak (*i.e.,* long window) is first set based on typical human heart rates (*i.e.,* between 50 and 200 beats per minute). Then, the long window is divided into four sub-windows. The maximum point of each sub-window is searched and considered as a possible candidate for R peaks. All candidates higher than the predefined threshold, $0.8 \times M_w$, are determined to be R peaks, where M_w denotes the maximum value in the long window. Finally, a single period of the ECG wave is extracted from the signal. The detection accuracy of the R peak is 96.6%, which is somewhat lower than that of state-of-the-art algorithms [12], but it does not severely affect compression performance because the main purpose of R peak detection is to segment a basic unit.



(a)

(b)

Fig. 1. Framework of the proposed algorithm: (a) encoder and (b) decoder.

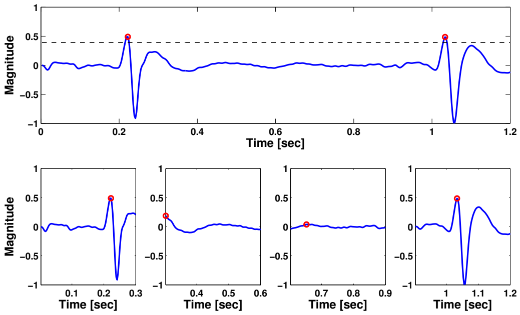


Fig. 2. Example of maximum searching algorithm, long and sub-windows.

B. Template analysis module

In the template analysis module, the extracted single beat is compared with the templates in the registered codebook. For each period, the residual signal $e(n)$ is obtained by subtracting the matched template from the input ECG as follows:

$$e(n) = x_p(n) - t_m(n), \qquad (1)$$

where $x_p(n)$ and $t_m(n)$ denote the pre-processed single ECG wave and the matched template, respectively.

The first extracted single period is registered to the template codebook because there is no registered template at the initial stage. Otherwise, the input signal is interpolated by using cubic spline method and similarities are then measured for each template in the codebook. The similarity between the pre-processed signal $x_p(n)$ and the matched template $\hat{t}(n)$ is measured by the NCC as follows:

$$C_{x\hat{t}} = \frac{\sum_n^N (x_p(n) - \mu_{x_p}) \sum_n^N (t_m(n) - \mu_{t_m})}{\sigma_{x_p}\sigma_{t_m}}, \quad (2)$$

where N denotes the length of the template, *e.g.,* 256, and μ and σ denote the mean and variance of each signal, respectively. Since the length of the extracted single period varies depending on the RR interval, it is interpolated to equalize its length to the templates before the NCC is calculated. If all the NCC values are lower than the preset threshold, *e.g.,* 0.9, the input ECG wave is registered to the template codebook as a new template. The maximum number of templates needs to be determined beforehand. In our experiments, it was enough to set the number 64. Otherwise, the template that has the maximum NCC is selected. After determining the matched template, an index of the matched template and a length of the input ECG wave are converted into the bitstream.

C. Residual coding module

To accurately reconstruct the ECG signals, the residual signal is encoded using a wavelet-based transform coding. Fig. 3 (a) shows a flow chart of the algorithm used to determine the number of coefficients while constraining the maximum error from the reconstructed signal with metric of percentage root mean difference (PRD). During this process, this algorithm increases the number of target DWT coefficients until the reconstruction error of the current period (*i.e.,* PRD$_{tmp}$) becomes smaller than PRD$_{max}$, which denotes the maximum PRD allowed in the current encoding step. The minimum number of coefficients that satisfies the target condition, namely, that PRD$_{tmp}$ is smaller than PRD$_{max}$ is set to the number of target DWT coefficients.

The target coefficients are selected by referencing the importance factor of the template mask. Fig. 3 (b) depicts the process used to extract a template mask from the matched template obtained at the previous layer. It is first pre-processed by removing the mean value and applying the squaring operator, then windowed to amplify the weight of medically important regions such as P peak, QRS complex, and T peak. In other words, since the selected templates are available in both the encoding and the decoding steps, we do not need to encode the position information separately.

After selection of the DWT coefficients, they are converted into the bitstream using a quantization process. The order and maximum/minimum (max/min) values of the target coefficients are found and quantized. The order of coefficients is encoded by "O_ϵ" that denotes the maximal range of coefficients with an exponential of two such as:

$$O_\epsilon = floor(\log_2(\max|\epsilon|)) + 1, \quad (3)$$

where ϵ denotes the target coefficients. The min/max values of the coefficients are first uniformly quantized from -2^{O_ϵ} to 2^{O_ϵ}. Coefficients are normalized to ϵ_n by the min-max normalization and uniformly quantized as follows:

$$\epsilon_q = floor(\epsilon_n / Q_{lv}), \quad (4)$$

where ϵ_q and Q_{lv} denote the quantized index and the quantization level, respectively.

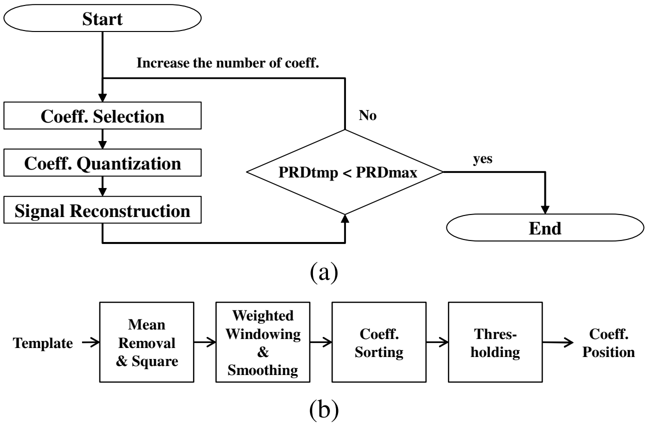


Fig. 3. Residual coding process: (a) determination of the number of target coefficients and (b) template masking.

TABLE I
BITSTREAM CONFIGURATION OF A SINGLE PERIOD ECG

Template-related	**Number of bits**
Template Index	6
Length of period	10
Residual-related	**Number of bits**
Number of coeff.	10
Order of coeff.	4
Maximum of coeff.	8
Minimum of coeff.	8
Quantized coeff.	Q_{bit}×(Number of coeff.)

Table I summarizes the configuration of the bitstream for the purpose of encoding a single period of the ECG wave, where Q_{bit} denotes the number of bits to represent a single DWT coefficient. The template-related bitstream includes template index and the length of the period. The residual-related bitstream includes the number of target coefficients, order/max/min of the coefficients and index of the quantized coefficients.

IV. EXPERIMENTS

The MIT-BIH arrhythmia database [13] is used to evaluate the performance of the proposed system. It consists of ECG waves and header files that contain patient information. The sampling rate of the signal is 360 Hz, and each sample is represented by 11 bits. The performance of the proposed coding algorithm is evaluated by CR and PRD as follows:

$$CR = \frac{N_o}{N_c}, \quad (5)$$

$$PRD = \sqrt{\frac{\sum_n^N (x_p(n) - x_r(n))^2}{\sum_n^N x_p^2(n)}} \times 100(\%), \quad (6)$$

where $x_r(n)$ represents the reconstructed signal, N_o and N_c denote the number of bits used to represent the original and the compressed signal, respectively. N_c includes all

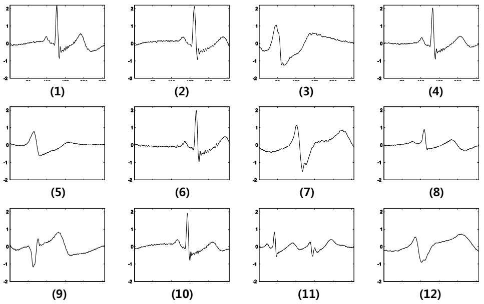


Fig. 4. An example of extracted templates (MIT-BIH record “106”).

the information required to represent registered codebooks and the encoded bitstream for ECG signals. Experiments are performed in two circumstances. The first involves an entire coding scheme including residual coding. The second involves compression without residual coding, which can obtain the highest compression performance.

Fig. 4 is an example of templates extracted from the record “106” who has a normal sinus rhythm and premature ventricular contraction (PVC). Some of templates are similar but the different positions of P, T peak result in a new template registration. From the histogram represented in Fig. 5, templates such as (1) and (2), which are normal sinus ECG waves, are the representative template for this patient. while other templates such as (3), (5), and (7) represent the PVC beats. The percentage of normal beats is 71.98%. By checking out the time index of the templates in (9), which is inverted QRS complex and (11), which is ventricular bigeminy, it is possible to detect the time at which arrhythmias occur.

Table II summarizes the performance of the CR under the different values of PRD$_{max}$ constraints. As the constraint PRD$_{max}$ becomes smaller, the CR also decreases, since a greater number of DWT coefficients needs be quantized. Fig. 6 depicts the distribution of the CRs obtained from the processing of all 48 records. In each box, the center line denotes median, the lower and upper edges represent the 25% and 75% positions, and the crossed marks denote outliers. Without including bits for residual coding, CR is higher than 40, but details of waves is not represented. Although the compression performance of the proposed algorithm is not much higher than that of the conventional approaches [7][9], the proposed algorithm has an advantage in terms of diagnosis because various types of arrhythmia can be detected directly by the distribution of coding parameters.

V. CONCLUSION

In this paper, a constrained two-layer compression technique was introduced, which has advantages in terms of extracting diagnostic information from the coded parameters. By integrating template-based and transform-based coding into two layers, the proposed algorithm not only extracted distinct templates that could be used in diagnostic applications but also provided a structure for scalable compression. Further study is necessary regarding the classification of cardiovascular diseases using the statistical distribution of templates.

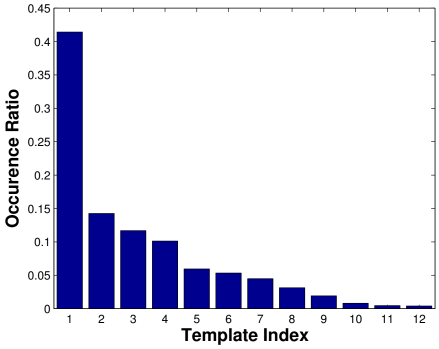


Fig. 5. Histogram of templates (MIT-BIH record “106”).

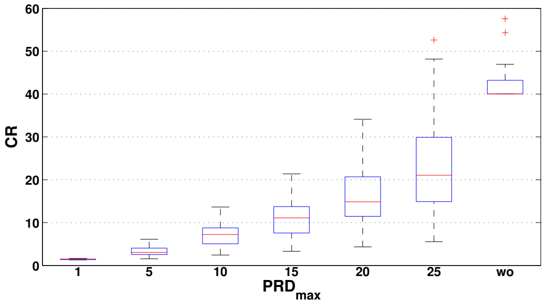


Fig. 6. Distribution of compression ratio (CR) at different PRD$_{max}$ options using 48 samples in the MIT-BIH database.

TABLE II

CR OF PROPOSED ALGORITHM

PRD$_{max}$	1	5	10	15	20	25	w/o
Mean	1.44	3.37	7.18	10.87	15.81	22.73	44.87
Median	1.42	3.05	7.22	11.10	14.86	21.05	40.02

REFERENCES

[1] J. R. Cox, F. M. Nolle, H. A. Fozzard, and G. C. Oliver, “AZTEC, a preprocessing program for real-time ECG rhythm analysis,” *IEEE Trans. Biomed. Eng.*, vol. BME-15, pp. 128-129, Apr. 1968.

[2] W. C. Mueller, “Arrhythmia detection program for an ambulatory ECG monitor,” *Biomed. Sci. Instrument.*, vol. 14, pp. 81-85, 1978.

[3] J . P. Abenstein and W. J . Tompkins, “New data-reduction algorithm for real-time ECG analysis,” *IEEE Trans. Biomed. Eng.*, vol. BME-29, pp. 43-48, Jan. 1982.

[4] S. M. S. Jalaleddine, C. G. Hutchens, R. D. Strattan, and W. A. Coberly, “ECG data compression techniques-A unified approach,” *IEEE Trans. Biomed. Eng.*, vol. 37, pp.329-343, 1990.

[5] A. S. Krishnakumar, J. L. Karpowicz, N. Belic, D. H. Singer, and J. M. Jenkins, “Microprocessor-based data compression scheme for enhanced digital transmission of Holter recordings,” *Computers Cardiol.*, Long Beach, CA, pp.435-437, 1980.

[6] G. Nave and A. Cohen, “ECG compression using long-term prediction,” *IEEE Trans. Biomed. Eng.*, vol.40, no.9, pp.877-885, 1993.

[7] S. B. Cho, Y. D. Lee, D. U. Jeong, and G. H. Hwang, “Implementation of novel ECG compression algorithm using template matching,” *in Proc. ICCCT*, pp.305-308, Dec. 2012

[8] E. Z. Ayari, R. Tielert, and N. Wehn, “Template-based Compression of ECG Signals”, *in Proc. EMBS*, pp.283-286, Aug. 2008.

[9] Z. Lu , D. Y. Kim and W. A. Pearlman “Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm,” *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp.849-856, 2000.

[10] N. Ahmed, P. J. Milne, and S. G. Hams, “Electrocardiographic data compression via orthogonal transforms,” *IEEE Trans. Biomed. Eng.*, vol. BME-22, pp. 484-487, Nov. 1975.

[11] R. Jane, P. Laguna, N. V. Thakor, and P. Caminal, “Adaptive baseline wander removal in the ECG: Comparative analysis with cubic spline technique,” *in Proc. Comput. Cardiol.*, pp. 143-146, 1992.

[12] J. Pan, and W. J. Tomkins, “A real-time QRS detection algorithm,” *IEEE Trans. Biomed. Eng.*, BME-32, pp. 230-236, 1985.

[13] G. B. Moody, and R. G. Mark, “The impact of the MIT-BIH Arrhythmia Database,” *IEEE Eng. in Med. and Biol.*, 20(3):45-50, (May-June), 2001.