

# A Fast Method for Abrupt Change Detection from Large-Scale Electrocardiogram (ECG) Time Series

Jin-Peng Qi<sup>1</sup>, Qing Zhang<sup>3</sup>, Jie Qi<sup>1</sup>, and Ying Zhu<sup>2</sup>

<sup>1</sup> College of Information Science & Technology, Donghua University,  
Shanghai 201620, P.R. China

{qipengkai, jieqi}@dhu.edu.cn

<sup>2</sup> Hunter New England Health, Royal North Shore Hospital, New South Wales, Australia  
ying.zhu@hnehealth.nsw.gov.au

<sup>3</sup> Australian e-Health Research Centre, CSIRO, UQ HSB 901/16, RBWH, Brisbane,  
QLD 4060, Australia  
qing.zhang@csiro.au

**Abstract.** In previous work, we proposed a promising method, named HWBST, for Change-Point (CP) detection from time series. However, the performance of HWBST is affected partly by the search criteria in terms of Binary Search Tree (BST). In this paper, we propose an improved method for fast CP detection from large-scale ECG time series, based on multi-level Haar Wavelet and Ternary Search Tree (HWTST, for short). In this method, we construct a ternary search tree termed TSTcD from a diagnosed time series by using multi-level HW. Then, we implement fast detection abrupt change from root to leaf nodes in TSTcD, by introducing two search criteria in terms of the data fluctuation in the left, right, and virtual middle branches of TSTcD. Based on the assembled and abnormal ECG samples, we evaluate the proposed HWTST by comparing with HWBST, KS, and T methods. The results show that the proposed HWTST is a faster and more efficient than HWBST, KS and T in terms of the computation time, error, accuracy, and distance of e.c.d.f.

**Keywords:** CP detection, ECG Time series, Large-Scale, Haar Wavelet (HW), Ternary Search Tree (TST).

## 1 Introduction

Currently, Change-Point (CP) detection has attracted considerable attention in the fields of data mining and statistics. CP detection has been widely studied in many real-world problems, such as atmospheric and financial analysis [1], intrusion detection in computer network [2], signal segmentation in data stream [3], as well as fault detection in engineering system [2, 4]. In the past decade, Wavelet Transform (WT), and its revised approaches have emerged as an important mathematical tool for analyzing time series [5-10]. WT is a promising approach for CP detection. It has found applications in anomaly detection, time series prediction, image processing, and noise reduction [6, 10-12]. WT can represent general functions at different scales and positions in a versatile and sophisticated manner, so the data distribution features can

be easily extracted from different time or space scales [12, 13]. As a simple WT, Haar Wavelet (HW), owns some attracting features including fast for implementation and able to analyze the local feature. HW is very useful to find discontinuity and high frequency change in time series, so it is a potential candidate in modern electrical and computer engineering applications, such as signal and image compression, as well as abnormality detection from time series [14, 15].

Previously, we proposed a HWBST method for fast CP detection [16]. HWBST is a promising method for detecting abrupt change from large-scale time series. However, the performance of HWBST is affected partly by binary search criteria, especially when abrupt change occurs near the middle boundary of BST. To resolve this problem, in this paper, we propose an improved method for fast CP detection from large-scale time series, named HWTST, based on multi-level HW and ternary search tree. In the proposed method, a ternary search tree (TST), named TSTcD is built based on a binary search tree, TcD derived from a diagnosed time series, by adding a virtual middle sub-tree; and then abrupt change is detected from TSTcD in terms of two search criteria. To evaluate the proposed method, we apply the HWTST, HWBST, KS, and T to detect abrupt change from both the assembled and the abnormal ECG samples. The detailed HWTST method is implemented as follows.

## 2 Method

### 2.1 Definition and Assumption [16]

Suppose  $Z = \{z_1, \dots, z_N\}$  is a diagnosed time series, we observe

$$Z = f(i/n) + X, i = 1, \dots, N, \tag{1}$$

where  $X = \{x_i\}_{i=1, \dots, N}$  are discrete and centred i.i.d. random variables, and  $f$  is a noisy mean signal with unknown distribution.

**Definition 1:** We define the data fluctuation between two adjacent segments in  $Z$  as,

$$D_{mn}(c) = \sup_{1 \leq L, R \leq N} \left| \frac{1}{m} \sum_{L=a}^c z_L - \frac{1}{n} \sum_{R=c+1}^b z_R \right|, \tag{2}$$

where  $Z = \{Z_L, Z_R\}$ ,  $Z_L = \{z_a, \dots, z_c\}$ ,  $Z_R = \{z_{c+1}, \dots, z_b\}$ , and  $m = c - a$ ,  $n = b - c - 1$ , and  $m + n \leq N$ . If a change point  $c$  occurs in  $Z$ , there exists a value  $c$  satisfies  $D_{mn}(c) > \delta$ ,  $z_c \in [z_1, z_N]$ , and  $\delta \in R$ .

**Hypotheses:**

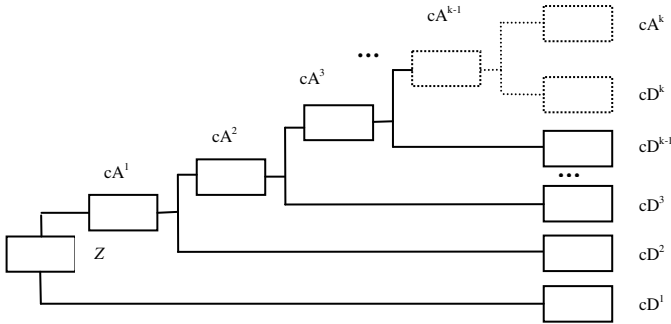
- (1)  $H_0$ : if  $D_{mn}(c) \leq \delta$ , no abrupt change occurs in  $Z$ .
- (2)  $H_1$ : if  $D_{mn}(c) > \delta$ , abrupt change occurs in  $Z$ .

where  $\delta \in R$  is a threshold of data fluctuation in a time series  $Z$  within an identical distribution. We wish to test  $H_0$  against  $H_1$  from observations. Thereafter, we assume that the number, the location, and the size of the function  $f$  are unknown. However, the upper bound of data fluctuation  $\delta$  is supposed to be known.

### 2.2 Multi-level HW

Generally, as shown in Fig.1, by using multi-level HW, a discrete time-series signal  $Z=\{z_1, z_2, \dots, z_N\}$ , can be decomposed into the  $k^{\text{th}}$ -level trend  $cA^k$ , and  $k$  level fluctuations, *i.e.*,  $cD^1, cD^2, \dots, cD^k, k=1, 2, \dots, \log_2 N$ . The  $k$ -level HW is the mapping  $H_k$  defined by [8],

$$Z \xrightarrow{H_k} (cA^k | cD^k | cD^{k-1} | \dots | cD^2 | cD^1), \tag{3}$$



**Fig. 1.** The diagram of multi-level HW for a time-series signal  $Z$ , it is composed of  $k$ -level  $cA$  and  $cD$  vectors, *i.e.*, the average and difference coefficients vectors

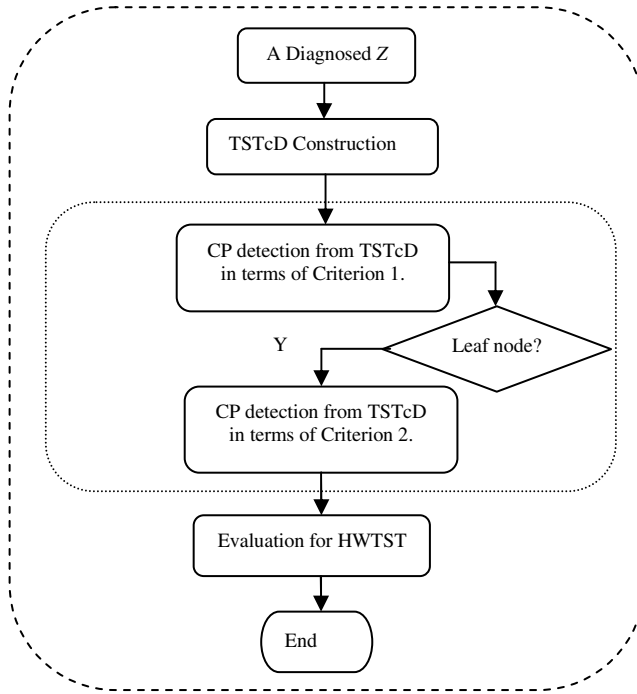
Suppose a diagnosed sample  $Z=\{z_1, z_2, \dots, z_N\}$  is decomposed by means of multi-level HW, we can represent the approximation and detail coefficient vectors by the following matrices, namely  $McA$  and  $McD$ :

$$McA = \begin{bmatrix} cA^0 \\ cA^1 \\ \dots \\ cA^k \\ \dots \\ cA^M \end{bmatrix} = \begin{bmatrix} cA_{0,1} & \dots & \dots & \dots & \dots & cA_{0,N} \\ cA_{1,1} & \dots & \dots & \dots & cA_{1,N/2} & \\ \dots & \dots & \dots & \dots & & \\ cA_{k,1} & \dots & cA_{k,j} & & & \\ \dots & \dots & & & & \\ cA_{M,1} & & & & & \end{bmatrix} \quad McD = \begin{bmatrix} cD^0 \\ cD^1 \\ \dots \\ cD^k \\ \dots \\ cD^M \end{bmatrix} = \begin{bmatrix} cD_{0,1} & \dots & \dots & \dots & \dots & cD_{0,N} \\ cD_{1,1} & \dots & \dots & \dots & cD_{1,N/2} & \\ \dots & \dots & \dots & \dots & & \\ cD_{k,1} & \dots & cD_{k,j} & & & \\ \dots & \dots & & & & \\ cD_{M,1} & & & & & \end{bmatrix}, \tag{4}$$

where  $cA^0 = cD^0 = Z = \{z_1, z_2, \dots, z_N\}$ ,  $0 \leq k \leq M = \log_2 N$ , and  $1 \leq j \leq N/2^k$ .

### 2.3 Overview of HWTST Method

The scheme of integrated HWTST method (Fig.2) is composed of three parts. First, a ternary search tree, TSTcD, is constructed from a diagnosed time-series  $Z$  by multi-level HW method. Second, abrupt CP is detected from root to leaf nodes in TSTcD, in terms of two search criteria. Last, the proposed method is evaluated by comparing with HWBST, KS, and T methods.



**Fig. 2.** The scheme of the integrated HWTST framework, it includes three parts: TSTcD construction, CP detection from TSTcD in terms of two ternary search criteria, and evaluation for HWTST method

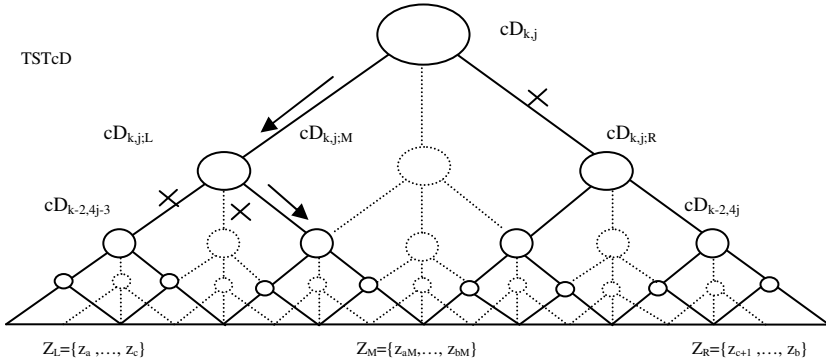
### 2.4 Construction of TSTcD

First, we build a binary tree TcD, in terms of McD. As shown in Fig.3, we deal the element  $cD_{k,j}$  in McD as the root, the other elements as non-leaf left and right child nodes, and the original  $Z = \{z_1, z_2, \dots, z_N\}$  as  $N$  leaf nodes, respectively. Then, a virtual middle-child node termed  $cD_{k,j;M}$  is added into non-leaf parent node  $cD_{k,j}$  in TSTcA.

### 2.5 CP Detection from TSTcD

**Definition 2:** Suppose the current non-leaf node  $cD_{k,j}$  is selected in TSTcD, with its left, virtual middle, and right child non-leaf node, namely  $cD_{k,j;L}$ ,  $cD_{k,j;M}$ , and  $cD_{k,j;R}$ , respectively. To measure the data fluctuation in different segments of  $Z$ , three statistic variables,  $D_{k,j;L}$ ,  $D_{k,j;M}$ , and  $D_{k,j;R}$  are defined as:

$$D_{k,j;L} = |cD_{k,j;L}| = \left| \frac{1}{(\sqrt{2})^{k-1}} \left( \sum_{L=a}^{aM-1} z_L - \sum_{R=aM}^c z_R \right) \right|, \tag{16}$$



**Fig. 3.** The scheme of TSTcD construction, the virtual non-leaf middle-child node is added

$$D_{k,j,R} = |cD_{k,j,R}| = \left| \frac{1}{(\sqrt{2})^{k-1}} \left( \sum_{L=c+1}^{bM} z_L - \sum_{R=bM+1}^b z_R \right) \right|, \tag{17}$$

$$D_{k,j,M} = |cD_{k,j,M}| = \left| \frac{1}{(\sqrt{2})^{k-1}} \left( \sum_{L=aM}^c z_L - \sum_{R=c+1}^{bM} z_R \right) \right|, \tag{18}$$

where  $a=2^k(j-1)+1, b=2^k * j$ , and  $c=2^k(j-1)+2^{(k-1)}$ ;  $aM=2^{k-1}(2j-2)+2^{k-2}+1$ , and  $bM=2^{k-1}(2j-1)+2^{k-2}$ ;  $1 \leq j \leq N/2^k$ , and  $2 \leq k \leq \log_2 N$ . Thereafter, we can introduce the first search criterion in terms of the data fluctuation defined above.

**Criterion 1:** Suppose the current non-leaf node  $cD_{k,j}$  is selected in TSTcD,  $2 \leq k \leq \log_2 N$ ,

- (a) If  $(\max(D_{k,j,L}, D_{k,j,R}, D_{k,j,M}) = D_{k,j,L})$  holds true, then the left-child node  $cD_{k,j,L}$  in TSTcD, is selected to be involved into the current search path;
- (b) If  $(\max(D_{k,j,L}, D_{k,j,R}, D_{k,j,M}) = D_{k,j,R})$  holds true, then the right-child node  $cD_{k,j,R}$  in TSTcD, is selected to be involved into the current search path;
- (c) If  $(\max(D_{k,j,L}, D_{k,j,R}, D_{k,j,M}) = D_{k,j,M})$  holds true, then the virtual middle-child node  $cD_{k,j,M}$  in TSTcD, is selected to be involved into the current search path.

In addition, we introduce another criterion to deal with the last leaf node in TSTcD, and then decide whether the selected leaf node is the potential CP or not. We define the second search criterion in terms of the data fluctuation in the last leaf node level as follows.

**Criterion 2:** Suppose the current selected non-leaf node is  $cD_{k,j}$ ,  $k=1$ , with left, and right-child leaf node,  $cD_{0,2j-1}$ , and  $cD_{0,2j}$ , namely,  $z_{2j-1}$ , and  $z_{2j}$ ,

- (a) If  $(|cD_{k,j}| > \sqrt{2}\alpha)$  is satisfied, then the right-child leaf node  $cD_{0,2j}$ , i.e.,  $z_{2j}$  is selected, and dealt as the estimated CP in  $Z$ ;
- (b) If  $(|cD_{k,j}| > \sqrt{2}\alpha)$  is satisfied, then there is no abrupt change detected from  $Z$ .

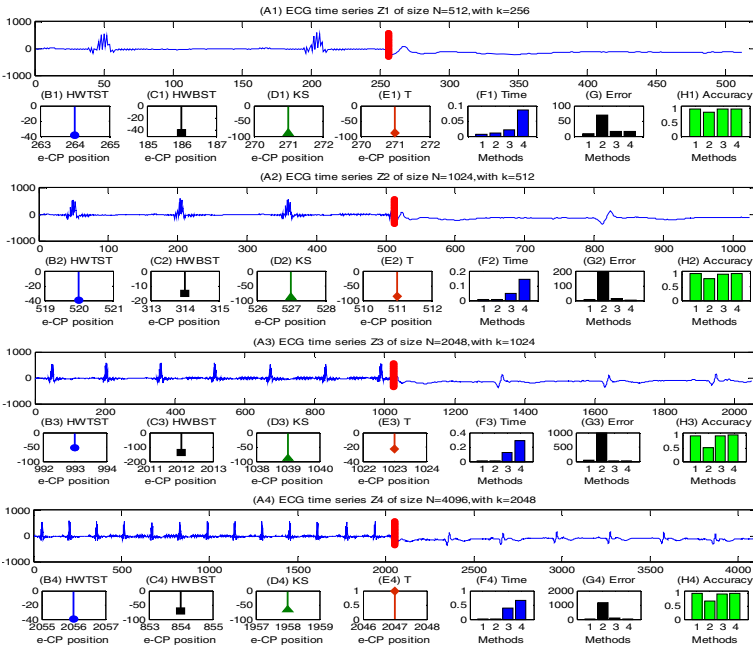
Based on two search criteria above, a fast CP detection is implemented, and then an estimated CP can be detected from root to leaf-nodes in TSTcD.

### 3 Results

To verify the performance of the proposed method further, we apply HWKS, and KS, HW, and T methods, to detect abrupt change from ECG time series provided by PhysioBank. In ECG experiments, we design the diagnosed ECG samples from different ECG datasets, including the MIT-BIH Normal Sinus Rhythm Database (NSRDB) [17], MIT-BIH Noise Stress Test Database (NSTDB) [18], and MIT-BIH Malignant Ventricular Arrhythmia Database (MVADB) [19].

#### 3.1 CP Detection from Assembled ECG Samples

First, we select a normal ECG dataset, 16265m from NSRDB, and an abnormal ECG dataset, 118e00m from NSTDB. Specifically, we take the normal ECG segment of size  $m$  as  $X_m$ , and the abnormal segment of size  $n$  as  $Y_n$ , respectively, and then assemble the diagnosed ECG sample  $Z=\{X_m, Y_n\}=\{x_1, \dots, x_m, y_1, \dots, y_n\}$ . Then, a single CP test



**Fig. 4.** The results of CP detection from the assembled ECG time series Z1-Z4. (A1)-(A4) the assembled Z1-Z4; (B1)-(B4),(C1)-(C4) ,(D1)-(D4), and (E1)-(E4) the e-CP detected from Z1-Z4, by HWTST, HWBST, KS, and T, respectively; (F1)-(F4) the diagram analysis for the computation time, (G1)-(G4) the error of e-CP, and (H1)-(H4) the accuracy for HWTST, HWBST, KS, and T, respectively. In histograms (F)-(H), '1', '2', '3', and '4' stands for HWTST, HWBST, KS, and T, respectively.

**Table 1.** The results of CP detection from the assembled ECG samples

M \ Z	Size, N	2^7	2^8	2^9	2^10	2^11	2^12	2^13	2^14	Mean
	CP, k	2^6	2^7	2^8	2^9	2^10	2^11	2^12	2^13	
Time	HWTST	.0050	.0060	.0060	.0060	.0060	.0060	.0070	.0090	.0063
	HWBST	.0060	.0060	.0070	.0080	.0080	.0100	.0140	.0240	.0103
	KS	.0100	.0130	.0200	.0440	.0119	.3910	1.420	5.431	.9176
	T	.0300	.0450	.0740	.1390	.2870	.6750	1.653	4.461	.9205
Error	HWTST	9	8	8	8	31	8	8	8	11
	T	63	63	15	1	1	1	1	1	18
	KS	15	15	15	15	15	90	52	70	36
	HWBST	62	4	70	198	988	1194	388	8074	1372
Accuracy	HWTST	.9297	.9688	.9844	.9922	.9849	.9980	.9990	.9995	.9821
	KS	.8828	.9414	.9707	.9854	.9927	.9780	.9937	.9957	.9675
	T	.5078	.7539	.9707	.9990	.9995	.9998	.9999	.9999	.9038
	HWBST	.5156	.9844	.8633	.8066	.5176	.7085	.9526	.5072	.7320

position is arranged at the middle part in each diagnosed ECG sample. The selected results are shown in Fig 4, and the results of the computation time, error and accuracy for all ECG samples of size from  $2^7$  to  $2^{14}$  are summarized in Table 1.

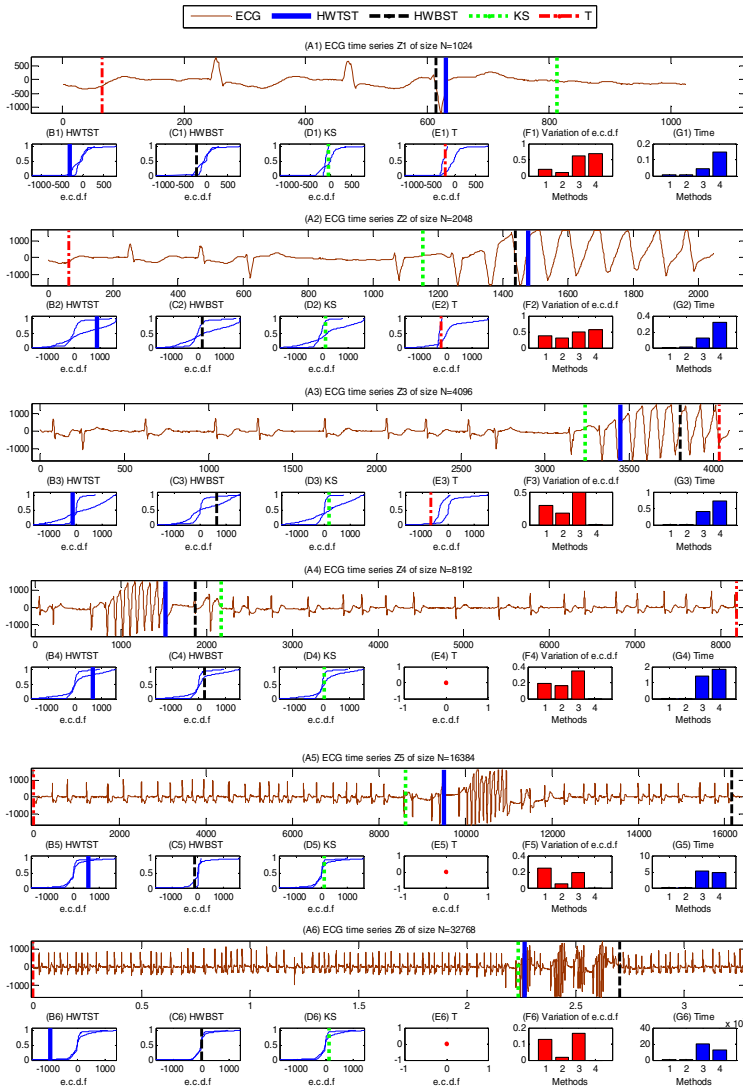
In this assembled ECG experiments, the results of CP detection illustrated in Fig.2, and Table 1 show that, the proposed HWTST can estimate CP position efficiently, and then distinguish the normal and abnormal segments from the assembled ECG samples with faster, smaller error and higher accuracy than HWBST, KS, and T. For HWBST, although faster than KS, and T, it has the biggest error and the lowest accuracy in all four methods. For KS, although slightly bigger error than T, it is faster than T, and has smaller error and higher accuracy than HWBST. For T, it has smaller error and higher accuracy than HWBST, whereas, it needs the longest computation time in all methods. These results indicate the proposed HWTST is more sensitive and efficient than HWBST, for CP detection at the middle boundary. Especially, it has the shortest computation time, the smallest error, and the highest accuracy in all four methods.

### 3.2 CP Detection from Abnormal ECG Time Series

To verify the performance of the proposed method further, we apply HWTST, HWBST, KS, and T to analyze the abnormal ECG time series directly. In this part, we select the abnormal ECG segment from 615m in the MVADB, i.e.,  $Z=\{Y_n\}=\{y_1, \dots, y_n\}$ , as a diagnosed ECG sample. To some extent, the distance of e.c.d.f can partly reflect the statistic fluctuation. Therefore, we take this variable as an indicator of the data fluctuation between two adjacent ECG segments divided by the estimated CP (e-CP) position. The selected results of CP detection are plotted in Fig.5, and the results for all abnormal ECG samples, including the e-CP position, computation time, and the variance of e.c.d.f are summarized in Table 2.

For abnormal ECG samples Z1-Z6 of size  $N$  from  $2^{10}$  to  $2^{15}$ , the results show that the proposed HWTST can detect the abrupt change position, and then divide the original ECG sample into two adjacent parts, with the shortest computation time in four methods; and it has bigger distance of e.c.d.f than HWBST, and T methods. For HWBST, although faster than KS, and T, it has the smallest variance of e.c.d.f in all four methods. For KS, it can detect CP with the maximal distance of e.c.d.f, whereas, it needs longer computation time than HWTST, HWBST, especially when ECG sample size is bigger. For T, it needs the longest averaged computation time in all four methods, although it has bigger variance of e.c.d.f than HWBST. These results

show that HWTST can capture abrupt change position from the diagnosed ECG samples more quickly and efficiently than HWBST, KS, and T. On the other hand, the detected CP is very useful to find a critical time from the diagnosed ECG time series, where a patient might encounter an important conversion between two different states



**Fig. 5.** The results of CP detection from abnormal ECG time series of size for  $2^{10}$  to  $2^{15}$ . (A1)-(A6) the abnormal ECG sample Z1-Z6; (B1)-(B6), (C1)-(C6), (D1)-(D6), (E1)-(E6) the e.c.d.f derived from two adjacent ECG segments in Z1-Z6, by HWTST, HWBST, KS, and T, respectively; (F1)-(F6) the diagram analysis of the distance of e.c.d.f, and (G1)-(G6) the computation time of HWKS, KS, HW, and T in Z1-Z6, respectively. In histograms (F)-(H), '1', '2', '3', and '4' stands for HWTST, HWBST, KS, and T, respectively.



**Table 2.** The results of CP detection from the abnormal ECG samples

Z	Size, N	2^10	2^11	2^12	2^13	2^14	2^15	Mean
e-CP	HWTST	630	1478	3448	1520	9503	22640	NA
	HWBST	614	1437	3804	1863	16167	27054	NA
	KS	813	1153	3237	2164	8610	22361	NA
	T	66	65	4036	8190	1	1	NA
Time	HWTST	.0060	.0060	.0070	.0060	.0090	.0140	.0080
	HWBST	.0080	.0090	.0100	.0140	.0200	.0350	.0160
	KS	.0430	.1320	.4100	1.409	5.294	20.46	4.625
	T	.1450	.3170	.7440	1.855	4.685	13.02	3.461
Variance of e.c.d.f	KS	.6260	.4919	.4991	.3456	.1909	.1632	.3861
	HWTST	.1944	.3750	.2916	.1941	.2475	.1297	.2387
	T	.6770	.5811	.0079	0	0	0	.2110
	HWBST	.0983	.3249	.1786	.1616	.0496	.0162	.1382

of health. Therefore, HWTST is very meaningful for inspecting and diagnosing different states of health from diagnosed ECG time series more quickly and efficiently.

### 4 Conclusion

In this paper, based on multi-level HW and ternary search tree, an improved method, named HWTST, is proposed for fast CP detection from large-scale ECG time series. In the proposed HWTST method, a ternary search tree, TSTcD, is built by adding a virtual middle-child sub-tree into TcD, which is derived from a diagnosed time series Z; and two search criteria are defined in terms of the data fluctuation in TSTcD. Then, an abrupt change is detected from root to leaf nodes of TSTcD in terms of two search criteria. Comparing with previous HWBST, KS, and T methods, the results show that the proposed HWTST is more efficient for fast detection CP from large-scale ECG time series, due to the shortest computation time, the smallest error, the highest accuracy in all four methods, and bigger variance of e.c.d.f than HWBST and T. In addition, HWTST has better performance than HWBST, especially when abrupt change occurs near the middle boundary.

**Acknowledgment.** I would like to thank my supervisor Prof. Qing. Zhang and Prof. Mohan Karunanithi in the Australia E-Health Research Centre for their assistances, supports and advices for this paper. This paper is supported by National Natural Science Foundation of China (No.61104154), and the Fundamental Research Funds for the Central Universities.

### References

1. Bolton, R.J., Hand, D.J.: Statistical fraud detection: A review. *Statistical Science*, 235–249 (2002)
2. Yamanishi, K., et al.: On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining and Knowledge Discovery* 8, 275–300 (2004)
3. Basseville, M.E., Nikiforov, I.V.: *Detection of abrupt changes: theory and application* (1993)

4. Murad, U., Pinkas, G.: Unsupervised profiling for identifying superimposed fraud. In: Żytkow, J.M., Rauch, J. (eds.) PKDD 1999. LNCS (LNAI), vol. 1704, pp. 251–261. Springer, Heidelberg (1999)
5. Gustafsson, F.: Adaptive filtering and change detection, vol. 1. Wiley, New York (2000)
6. Alarcon-Aquino, V., Barria, J.A.: Anomaly detection in communication networks using wavelets. *IEE Proceedings-Communications* 148, 355–362 (2001)
7. Khalil, M., Duchêne, J.: Detection and classification of multiple events in piecewise stationary signals: comparison between autoregressive and multiscale approaches. *Signal processing* 75, 239–251 (1999)
8. Kobayashi, M.: Wavelets and their applications in industry. *Nonlinear Analysis: Theory, Methods & Applications* 47, 1749–1760 (2001)
9. Percival, D.B., Walden, A.T.: Wavelet methods for time series analysis, vol. 4. Cambridge University Press (2006)
10. Salam, M., Mohamad, D.: Segmentation of Malay Syllables in connected digit speech using statistical approach. *International Journal of Computer Science and Security* 2, 23–33 (2008)
11. Qi, J., et al.: Kinetic theory approach to modeling of cellular repair mechanisms under genome stress. *PloS One* 6, e22228 (2011)
12. Tseng, V.S., et al.: Segmentation of time series by the clustering and genetic algorithms. In: Sixth IEEE International Conference on Data Mining Workshops, ICDM Workshops 2006, pp. 443–447 (2006)
13. Alarcon-Aquino, V., Barria, J.: Change detection in time series using the maximal overlap discrete wavelet transform. *Latin American Applied Research* 39, 145–152 (2009)
14. Darkhovski, B.S.: Nonparametric methods in change-point problems: A general approach and some concrete algorithms. *Lecture Notes-Monograph Series*, 99–107 (1994)
15. Walker, J.S.: A primer on wavelets and their scientific applications. CRC press (2002)
16. Qi, J.P., Zhang, Q., Qi, J.: A Fast Method for Change Point Detection from time series Based on Haar Wavelet and Binary Search Tree (HWBST). In: presented at the The 19th World Congress of the International Federation of Automatic Control Cape Town, South Africa (to be published)
17. Goldberger, A.L., et al.: Physiobank, physiotookit, and physionet components of a new research resource for complex physiologic signals. *Circulation* 101, e215–e220 (2000)
18. Moody, G.B., et al.: A noise stress test for arrhythmia detectors. *Computers in Cardiology* 11, 381–384 (1984)
19. SD, G.: Development and analysis of a ventricular fibrillation detector, M.S. thesis, MIT Dept. of Electrical Engineering and Computer Science (1986)