Personal Identification via Heartbeat Signal

Shi-Jinn Horng
Dept. of Computer Sci. and Information Engineering
National Taiwan University of Science and Technology, Taipei, Taiwan
Dept. of Computer Engineering
Dongguan Polytechnic, China
e-mail: horngsj@yahoo.com.tw

Xuan-Zi Hu and Bin Li
Dept. of Computer Engineering,
Dongguan Polytechnic, China
e-mail: 40436116@qq.com

Naixue Xiong
School of Computer Science and Technology, Tianjin University, Tianjin, China
email: xiongnaixue@gmail.com

***Abstract*—This study is trying to improve the biometric system using the heartbeat signal. The proposed algorithm calculates the contribution of all extracted features to biometric recognition. The efficiency of the proposed algorithms is demonstrated by the experiment results obtained from the Multilayer Perceptron, Naïve Bayes and Random Forest classifier applications based on the extracted features. The results were evaluated via the Multilayer Perceptron, Naïve Bayes and Random Forest classifier models; the true positive rates are then 94.6078%, 92.1569% and 90.3922%, respectively. Compared to existing methods, the proposed method has the better results.**

Keywords-PPG; biometric system; heartbeat signal;

I. INTRODUCTION

The biometric system plays an important role in network security issues. The uniqueness of each individual was needed to improve the security system applications. Instead of using password for access control, biometric identification can be used for authentication. Biometric information is hard to be duplicated, lost, forgotten, shared or transferred because it is a part of human body. Unfortunately, hackers can possibly get into system via counterfeit biometric information. Fingerprints can be affected by chemical reactions for the people who are working in the industry. Furthermore, biometric identification system using voice can be changed seriously due to aging and health condition. Finally, EEG and ECG-based methods are impractical as various electrodes are required in order to acquire the bio-signals.

In this study, the photoplethysmography (PPG) signal was used for data input that is capable of implementing identification functionality. The advantage of using the PPG signal is widely used; it is easier and more affordable price sensors.

The method proposed for biometric recognition in this study is composed by data acquisition, pre-processing, PPG signaling and the feature extraction of PPG signal using smoothing PPG signal and its first and second derivatives. The process of data acquisition was provided by 50 volunteers through a PPG data acquisition card. Existing approaches are reviewed in the following.

Gu et. al. [1] provided a new approach of human verification using the PPG signals acquired easily from the fingertips. For the group consisting of 17 healthy subjects, they performed experimental studies by obtaining four feature parameters from digitized PPG signals. A feature vector template was formulated using the recorded signals, and later on, the discriminant function was applied in order to verify the data. This promising method of human identification finally achieved a 90% success.

In Yao et. al. [2], two important conclusions like the derivatives of PPG signals, and the consistency of subjects within themselves and the distinguishability among different subject are examined. Data taken using Pulse Oximeter Sensor, statistically the results of same subject have a constant time interval against the generated maximum/minimum points, and the derivatives can certainly indicate the features of one's PPG signal and can be used as biometrics for recognition.

In Spachos et. al. [3], on the other hand, the feasibility of the application of PPG signal as a single biometrical feature along with the signal-processing methods for the matter involved is being researched. The PPG signals were acquired from the fingertips of 29 healthy subjects using BvpPLUX System from OpenSignal PPG Dataset and also using NONIN pulse oximeter from BioSec PPG Dataset. The classification was applied using the Nearest Neighbor and Majority Voting for the data to match the input signal. The accuracy results of identification depend on the dataset used. This can occur because of the influence of the circuit, the sensor and the current state of data collection. The experimental results suggesting biometrics for identification can be used when PPG signals come under a controlled environment with infallible sensors.

Wei et. al. [4] addressed that PPG signals could reflect numerous physiological parameters, such as heart functions, blood vascular elasticity and blood viscosity. This is a new non-invasive method with the advantages like smoothness and accuracy. It is important to find out efficient pre-processing

978-1-5386-9403-9/18/$31.00 ©2018 IEEE
DOI 10.1109/PAAP.2018.00034

and feature extraction algorithms in order to deal with the original PPG signal that could be affected by many other factors. Most of the practical methods include median and FIR (Finite Infinite Response) filtration. In this study, a new algorithm is recommended in order to eliminate the wavelet transform-based baseline deviation, like a sophisticated differential algorithm that created an effective platform for determining the physiological parameters.

Gu et. al. [5] had showed a fuzzy-logic approach to examine the feasibility of the application of PPG signals as a new method in the identification of humans. The PPG signals were acquired from the fingertips of 17 healthy subjects and were used as fuzzy entries for the classification of four distinctive features such as the peak number, the upward slope, the downward slope, and the time interval. This fuzzy-based decision-making can reach up to 94% in the same testing and 82.3% for two different trials. This result suggests that this new PPG-based biometry is potentially feasible in the verification of humans.

In Wan et. al. [6], the design of an amplifier circuit intended for extracting the DC component of the signal is being negotiated for PPG signals. Consequently, a high AC signal with SNR (signal-to-noise ratio) is acquired from a raw PPG signal, adding a bias-adjusted circuit to the amplifier. This hardware development resulted in acquiring a better signal quality and a data handling convenience in recognition (identification).

In Singh et. al. [7], the fingerprint of someone could be imitated by placing a thin film or using the artificial copy of that print in a biometric system operating via finger scanning. The uniqueness of a finger impact profile was approved in the preliminary studies. This study creates researches into the possibility to utilize the PPG signal as an additional parameter along with the fingerprint.

Kavsaoglu et. al. [8] got data using a microcontroller and sensors DCM03 then using k-NN (k-Nearest Neighbor) to do classification. Forty different features were used for feature extraction stage, including augmentation index, systolic and diastolic peak, pulse width, and peak-to-peak interval. When the results were evaluated for the k-NN classifier model created along with the proposed algorithm, an identification of 90.44% for the 1st configuration, 94.44% for the 2nd configuration, and 87.22% for the 3rd configuration has successfully been attained.This template provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. PLEASE DO NOT RE-ADJUST THESE MARGINS. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. Approach

A. Data Acquisition

In this study, PPG signals are acquired from a total of fifty one healthy volunteers, and twenty one of them are male and the remaining persons are female. The data are obtained from their right index fingers while they are seated in a calm position. Total 90-period-signal is acquired from each individual at two different time spans. 30 characteristic features are extracted for each period and these characteristic features are used for the purpose of classification.

B. Identification System of PPG Signals

The block diagram of the identification system using PPG signals is shown in Figure 1. In this system, the PPG signals are acquired by an Arduino and a pulse sensor with a 5 Hz sampling frequency. In order to debug the noises in the PPG signal as a pre-processing, a band-pass filter out of 3rd order Butterworth low-pass and high-pass filters with cutoff frequencies of .8Hz and 5Hz is utilized. Low order polynomial polyfit and the polynomial polyval are used to detrend the signal with obvious baseline drift. This drift is mainly caused by the breath signal and the motion artifact. A Polynomial method is proposed to eliminate the influence of the breath signal. The result of the polynomial reconstruction method is shown in Figure 2.

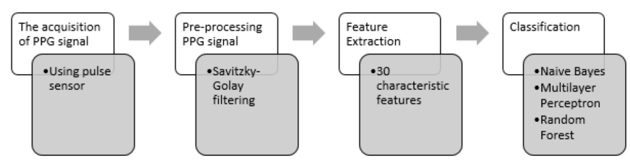
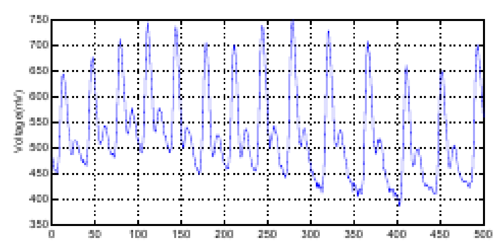
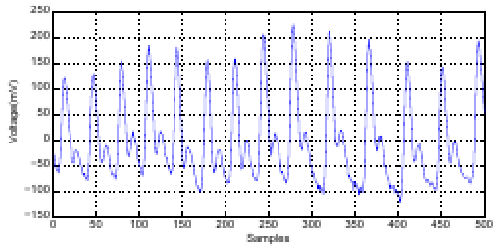


Figure 1: The block diagram of an identification system using Heartbeat PPG Signals



(a) PPG signal with a trend



(b) PPG signal after remove the trend

Figure 2: Detrend signal

The method for detecting the peaks is implemented in MATLAB®. Peaks detection is implemented using the function findpeaks, in which, the data of heartbeat have to be input and there are two input arguments have to be defined as shown in Figure 3. The MinPeakDistance defines the function used to specify the minimum peak distance, or minimum separation between peaks as a positive integer. We can use the MinPeakDistance option to specify the small peaks that occur in the neighborhood of a larger peak. When we specify a value for MinPeakDistance, the algorithm initially identifies all the peaks in the input data and sorts those peaks in descending order. Beginning with the largest peak, the algorithm ignores all identified peaks not separated by more than the value of MinPeakDistance. The MinPeakHeight function finds only those peaks that are greater than the value of MinPeakHeight. findpeaks only returns peaks that exceed the MinPeakHeight.

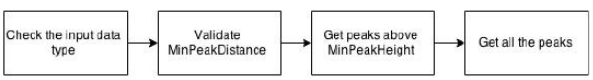
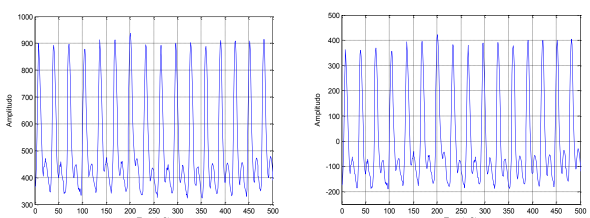
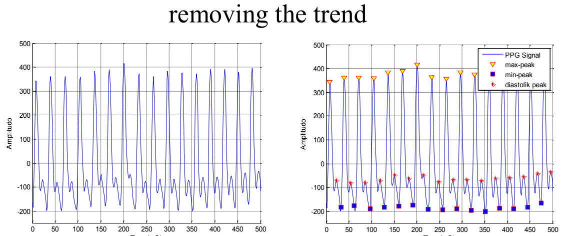


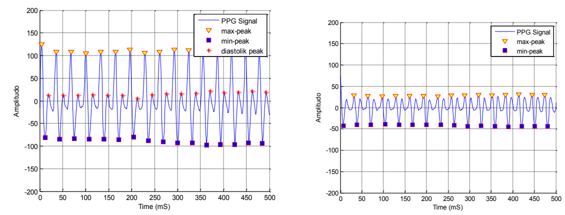
Figure 3: Block diagram of findpeaks function



(a) Original signal (b) PPG signal after removing the trend



(c) Savitzky-Golay Filtering in PPG signal (d) Peak detection in Smoothing PPG signal



(e) Peak detection in 1st derivative PPG signal (f) Peak detection in 2nd derivative PPG signal

Figure 4: Peak detection result

Figure 4(a) shows the original signal which results from the sensor with a baseline shift and therefore does not represent the true amplitude. In order to remove the trend, fit a low order polynomial to the signal and use the polynomial to remove the trend. Figure 4(b) shows the detrend PPG signal. Before peak detection step, determine the location of the each proposed peak. Thresholding the peaks to locate the proposed and required peaks can remove unwanted peaks caused by noise and it can be done by noise filtering techniques. In this paper, Savitzky-Golay filtering is used to remove the noise in the signal and the result is shown in Figure 4(c). After detrending signal and noise filtering as shown in Figure 4(b) and 4(c), find the main-complex peaks, which are the most prominent repeating peaks in the PPG signal, such as systolic peak (maximum value), diastolic peak and minimum peak. Figure 4(d) shows the obtained peaks. The derivative for one-dimensional signals can be calculated by Equations 1 and 2. The graphs of the 1st and 2nd derivatives are shown in Figures 4(e) and 4(f), respectively.

$$1^{st}\ derivative: \frac{\partial f}{\partial x} = f(x+1) - f(x) \quad (1)$$

$$2^{nd}\ derivative: \frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x) \quad (2)$$

In Figures 5, we do some labels on the PPG signal and its corresponding first derivative and second derivative. For example, x, y and z labeling from smoothing signal represent the systolic peak, the diastolic peak and the minimum peak with their corresponding times labeled as t1, t2, and t3, respectively. The other features can be calculated by these major feature spots detected in the time domain. The time between two systolic peaks is referred to as Time Peak to Peak (tpp). The distance between the beginning and the end of the PPG waveform of the ith cycle is labeled as Time Pulse Interval (tpi). The time between to diastolic peaks labeled as Time between Diastolic Peaks (tdp). Augmentation index (AI) is defined as a ratio calculated from the blood pressure waveform as the ratio of diastolic peak to systolic peak (diastolic peak/systolic peak). Takazawa et. al. [12] defined the augmentation index (*AI*) as the ratio of y to x as in Equation 3. Rubins et. *al.* [13] used the reflection index as in Equation 4 and introduced an alternative augmentation index. For other features such as ΛT1, ΛT2 and ΛT3 corresponds to time between diastolic and systolic peaks, time between minimum and systolic peaks and time between minimum and diastolic peaks, respectively. And the remaining features are t1/tpp, t2/tpp, t3/tpp and ΛT1/tpp is labeled from the smoothing PPG signal.

The initial peak point for the first derivative and second derivative are a1 and a2 respectively. It then comes b1 and e1 points for the first derivative and b2 for the second derivative, following the position of systolic peak point. Corresponding times of each feature from both first derivative and second derivative signals are labeled as a1 time, b1 time, c1 time, a2 time, and b2 time, respectively.

AI= diastolic peak(y)/systolic peak (x) (3)

Alternative AI= (systolic peak(x)-diastolic peak(y))/systolic peak(x) (4)

A total of 30 characteristic features are calculated. Table 1 shows all 30 features defined for the system. For the example, in the second column of Table 1, the value of each label from the second period signal of Figure 5 is shown.

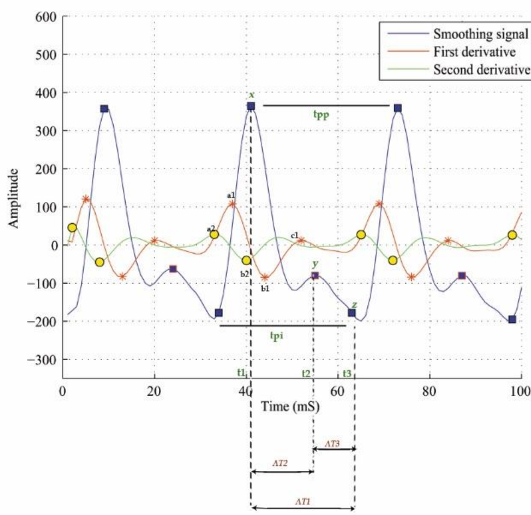


Figure 5: The specified parameters used to derive the characteristic features from the PPG signal.

Table 1: All 30 features defined for the system

No.	Features	The values of each label
1	Systolic peak (x)	363.907
2	Systolic peak time (t1)	41
3	Time peak to peak (tpp)	32
4	Minimum peak (z)	186.3494
5	Minimum peak time (t3)	63
6	Time Pulse interval (tpi)	35
7	Diastolic peak (y)	-78.5053
8	Diastolic peak time (t2)	55
9	Time between diastolic peaks (tdp)	32
10	y/x (augmentation index)	-0.21573
11	(x-z)/x (alternative augmentation index)	1.215729
12	t1/x (systolic peak output curve)	0.112666
13	y/(tpi-t3) (diastolic peak downward curve)	2.80376
14	t1/tpp	1.28125
15	t2/tpp	1.96875
16	t3/tpp	1.71875
17	ΛT1 (time between diastolic and systolic peaks)	14
18	ΛT2 (time between minimum and systolic peaks)	22
19	ΛT3 (time between minimum and diastolic peaks)	8
20	ΛT1/tpp	0.4375
21	a1	108.1641
22	a1 time	37
23	b1	84.50754
24	b1 time	44
25	c1	12.24569
26	c1 time	52
27	a2	28.34428
28	a2 time	33
29	b2	40.01237
30	b2 time	40

III. EXPERIMENTAL RESULTS

A. *Classification Result*

After processing each subject in the dataset, the extracted features are used as an input data in the classification stage. Four scenarios are used to test the accuracy of the system: training set, supplying test set, cross validation and percentage split. In the training set option, testing is performed by using the training data itself. In the cross-validation option, 10-fold is used. As for the percentage split option, from 90-period-signal, 66% is used for the training data and the rest is used for testing data. The data used for training and testing is chosen by the system itself. For the supplying test set option, from 90-period-signal taken, 70-period-signal of the initial data are used for training data and the remaining data are used for testing data. Furthermore, classification methods using Naïve Bayes, Multilayer Perceptron and Random Forest are proposed, respectively.

Table 2 the comparison of different classification methods to see how the performance varies across different architectures are shown. The performance results show accuracy (ACC), for the following architectures: Fuzzy, K-nn, Naïve Bayes, Random Forest and Multilayer Perceptron. The first column of Table 2 the performance for fuzzy logic by Gu et al. and the second column the performance for k-nn by Kasgaovlu et al. are shown for seventeen subjects and thirty subjects, respectively. In our study, fifty one subjects joined for the experiments, this method can achieve good performance for more data samples. In this case, a 0.6 % increase in the classification success is attained, compared with Fuzzy logic method and a 0.16 % increase compared with K-nn method.

Table 2 comparison with different classification methods

	Fuzzy	K-nn	Naïve Bayes	Random Forest	MLP

AC C	94%	94.44%	92.15	90.39%	94.6%

IV. Conclusions

This study has tested the ability of PPG signals for biometric identification system. Based on the research that has been done can be concluded that:

The designed system can identify the heartbeat of each individual. Feature extraction based on the three major peak value of the photoplethysmography signal.

The results were evaluated via the Multilayer Perceptron, Naïve Bayes and Random Forest classifier models; the true positive rates are then 94.6078%, 92.1569% and 90.3922%, respectively. The obtained results showed that the proposed algorithm and the biometric identification model based on this developed PPG signal are very promising for contact less recognizing systems.

Acknowledgment

This work was supported in part by the Ministry of Science and Technology under contract numbers 106- 2221-E-011 -149 -MY2 and 107-2218-E-011-008-, and it was also partially supported by the “Center for Cyber-physical System Innovation” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

[1] Y. Gu, Y. Zhang, and Y. Zhang, "A novel biometric approach in human verification by photoplethysmographic signals," in Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on, pp. 13-14, April 2003.

[2] J. Yao, X. Sun, and Y. Wan, "A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier," in Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pp. 4576-4579, Aug 2007.

[3] P. Spachos, J. Gao, and D. Hatzinakos, "Feasibility study of photoplethysmographic signals for biometric identification," in Digital Signal Processing (DSP), 2011 17th International Conference on, pp. 1-5, July 2011.

[4] C. Wei, L. Sheng, G. Lihua, C. Yuquan, and P. Min, "Study on conditioning and feature extraction algorithm of photoplethysmography signal for physiological parameters detection," in Image and Signal Processing (CISP), 2011 4th International Congress on, vol. 4, pp. 2194-2197, Oct 2011.

[5] Y. Gu and Y. Zhang, "Photoplethysmographic authentication through fuzzy logic," in Biomedical Engineering, 2003. IEEE EMBS Asian-Pacific Conference on, pp. 136-137, Oct 2003.

[6] Y. Wan, X. Sun, and J. Yao, "Design of a photoplethysmographic sensor for biometric identification," in Control, Automation and Systems, 2007. ICCAS '07. International Conference on, pp. 1897-1900, Oct 2007.

[7] Y. Singh and P. Gupta, "Correlation-based classification of heartbeats for individual identification," Soft Computing, vol. 15, no. 3, pp. 449-460, 2011.

[8] A. R. Kavsaoglu, K. Polat, and M. R. Bozkurt, "A novel feature ranking algorithm for biometric recognition with fPPGg signals," Computers in Biology and Medicine, vol. 49, no. 0, pp. 1 - 14, 2014.

[9] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[10] D. Setsirichok, T. Piroonratana, W. Wongseree, T. Usavanarong, N. Paulkhaolarn, C. Kanjanakorn, M. Sirikong, C. Limwongse, and N. Chaiyaratana, "Classification of complete blood count and haemoglobin typing data by a c4.5 decision tree, a nave bayes classifier and a multilayer perceptron for thalassaemia screening," Biomedical Signal Processing and Control, vol. 7, no. 2, pp. 202 - 212, 2012.

[11] M. Joel and G. Yury, http://pulsesensor.com/.

[12] K. Takazawa, N. Tanaka, M. Fujita, O. Matsuoka, T. Saiki, M. Aikawa, S. Tamura, and C. Ibukiyama, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform.," in Hypertension, p. 32(2): 365-370, August 1998.

[13] U. Rubins, A. Grabovskis, J. Grube, and I. Kukulis, "Photoplethysmography analysis of artery properties in patients with cardiovascular diseases," in 14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (A. Katashev, Y. Dekhtyar, and J. Spigulis, eds.), vol. 20 of IFMBE Proceedings, pp. 319-322, Springer Berlin Heidelberg, 2008.