Photoplethysmogram signal based Biometric Recognition using Linear Discriminant Classifier

Samik Chakraborty
AEIE Dept.
Heritage Institute of Technology
Samikasan0516@gmail.com

Saurabh Pal
Applied Physics Dept.
University of Calcutta
Spal76@gmail.com

***Abstract*— A preliminary study on photoplethysmogram (PPG) based biometry system is presented here. PPG is a physiological signal related to cardiac output and blood flow saturation in body. Recently it is reported that being an automatic physiological phenomenon PPG and other biosignals can be used as biometric parameters for human authentication. In this work, 12 number of features are extracted from filtered PPG and its derivatives and Linear Discriminant Analysis (LDA) is used for classification over the statistical parameters extracted from the feature set. 100% accuracy is achieved for 15 number of data captured using Biopac MP 45.**

Keywords*— Biometric authentication, PPG, standard deviation, Linear Discriminant Analysis

I. INTRODUCTION

Biometric system based on physical or behavioral characteristics of a human being is more reliable and more secure system than the traditional identification technology [1]. In literature several physiological and behavioral characteristics are used as biometric features such as, face [2], fingerprint [3], palmprint [4], palm vein [5], iris [6], ECG [7], EEG [8] and so on. These applications based on biometric approaches provide a promising and convincing future of human recognition. However, a fingerprint can be modified with latex, face recognition can be faked with an artificially disguised, audio recorder can be used for voice [9], and the EEG and ECG-based methods are to some extent complicated to acquired the bio-signals. Photoplethysmography (PPG) is a non-invasive electro-optical methodology which provides information concerning the amount of blood changes in the micro vascular bed of tissue at the skin surface initiated by the heart [10]. Compared with other biometric approaches, PPG based technique has many distinct advantages like low development cost, easy to use without any complicacy, and suitably accessible. Moreover, PPG signals offer a noninvasive and correct methodology to get valuable physiological information like heart rate, and blood flow [11] etc.

In literature, few approaches based on PPG are proposed so far. Gu *et al.* [12] offered a new approach of human verification using the PPG signals acquired simply from the finger-tips and obtained four feature parameters from digitized PPG signals of 17 healthy subjects and finally achieved 94% success on verification by applying Euclidean distance. In another study [13] a fuzzy-logic based approach is proposed to examine the feasibility of the application of PPG signals as a new method in the identification of humans. Yao *et al.* [14] suggested derivatives of PPG signals can certainly indicate the features of one's PPG signal and can be used as biometrics for recognition. Spachos *et al.* [15] demonstrated that PPG signals can be used as bio-measures for identification purposes if PPG signals are collected under controlled environment and with accurate sensors. Bonissi *et al.* [16] presented a preliminary study on continuous biometric authentication techniques based on PPG signals by using simple and fast correlation technique and also suggested that PPG signal can be used as biometric as sufficient discrminability present in it. Kavsaoğlu *et al.* [17] proposed a feature ranking algorithm for a sum of 40 time domain features, acquired from first and second derivatives of the PPG signal for biometric identification and calculates the contribution of each feature to biometric recognition. In another work, Jaafar *et al.* [18] proved the robustness and reliability of Acceleration Plethysmogram signal as a biometric recognition system by acquiring data from 10 subjects and achieved 97.5% identication rate.

In the present work, a preliminary study with 15 healthy subjects is made where the feasibility of using 1st and 2nd derivative based features of PPG for biometric analysis is tested using linear discriminant analysis based classification technique.

II. METHODOLGY

Data are collected from 15 healthy subjects of ages between 20 and 45 under normal relax state, by using BiopacMP 45 [19] data acquisition system consisting of an LED and a photo detector is attached on the fingertip (right index finger). PPG signals are recorded continuously for about three minutes from each person and converted into digital signals at the sampling rate of 1K samples. To minimize motion artifact, the subjects are encouraged to keep their fingers still.

The general block diagram of biometric recognition system based on PPG signal is shown in Fig. 1. PPG signal of a subject is collected, which is act as a raw data input. The raw data are preprocessed using a low pass FIR filter.

The Preprocessing stage removes or suppresses noise like power line interference from the raw input.

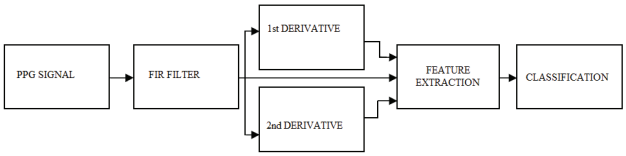

Fig. 1. Block diagram of proposed biometric recognition system

Next the first and the second derivative signals are calculated from the normalized filtered PPG signal. Unique features are extracted from the normalized filtered PPG and its derivatives. These features are used for the purpose of classification. The proposed algorithm is elaborated below.

A. Data Processing

In order to immune the noises in the PPG signal, the raw data are preprocessed by using a 2nd order low-pass Butterworth filter of 10 Hz cut-off frequency which is shown in Fig. 2.

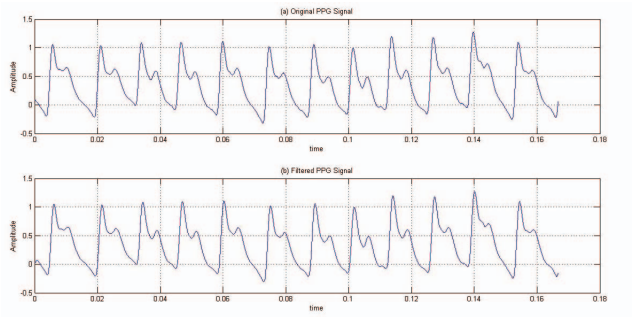

Fig. 2. a) Original PPG Signal, b) Filtered PPG Signal

B. Feature Extraction

After filtering, filtered PPG signal $f(n)$ is normalized in the interval [0, +1] by a standard method to obtain a shape response function $fnorm(n)$.

Next, the first and second derivatives of the PPG are calculated from the normalized filtered PPG signal ($fnorm(n)$) using the following formulas,

For first derivative,

$$f_1(n) = \frac{df_{norm}}{dn} = f_{norm}(n+1) - f_{norm}(n) \quad (1)$$

For second derivative,

$$f_2(n) = \frac{d^2 f_{norm}}{dn^2} = f_{norm}(n+1) + f_{norm}(n-1) - 2f_{norm}(n) \quad (2)$$

Then the points of maximum amplitudes A and B (Fig. 3a) as two consecutive peak points are detected. The starting point of these two consecutive peaks is detected by searching the closest to zero point just before *h* in the first derivative plot so that two complete pulses can be picked up by the data set. Pulse interval (t_{pi}) is calculated by measuring the time difference of starting points of two complete consecutive pulses and by measuring the time difference between systolic peaks of two consecutive pulses, peak to peak time (t_{pp}) is achieved.

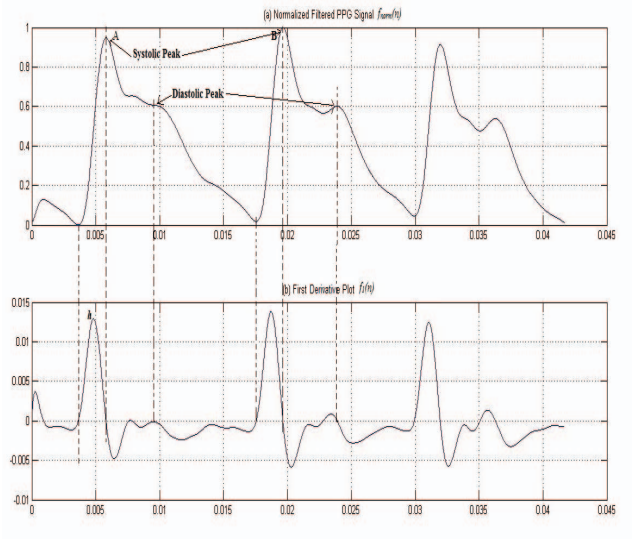

Fig. 3. a) Normalized filtered PPG signal b) 1st derivative PPG signal

However, the Dicrotic notch (***b***) and Diastolic peak (***c***) are detected by using second derivative PPG signal as shown in Fig. 4.

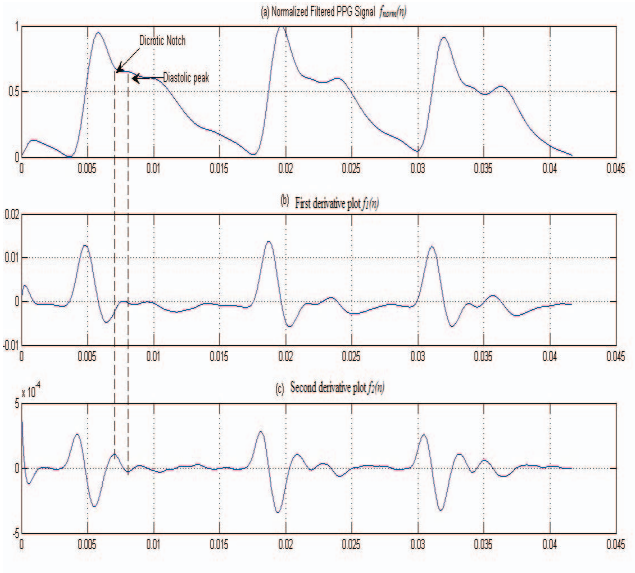

Fig. 4. a) Normalized filtered PPG signal b) 1st derivative PPG signal c) 2nd derivative PPG signal

In most cases, the largest minimum points within the second order derivative correspond to the systolic peaks and the maximum point following these typically corresponds to the dicrotic notch (***b***). The timing of the dicrotic notch (t_2) is located as the timing of the maximum point in second derivative plot following the systolic peak in each pulse cycle. After that, diastolic peak (***c***) is detected at the point of inflection within the second order derivative following the

dicrotic notch point and corresponding time is identified as diastolic time (t_3). The systolic-diastolic peak-to-peak time (ΔT) is then calculated.

After finding systolic and diastolic peak, augmentation index is defined as the ratio ***c*** to ***a***, whereas alternative augmentation index is defined as the ratio of (***a-c***) to ***a***. Next, pulse width (***d***) is detected at the semi height of the systolic peak. A typical normalized PPG signal is shown in Fig. 5.

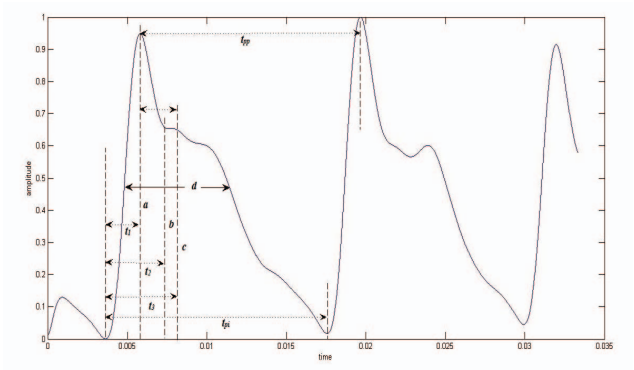

Fig. 5. Characteristic features acquired from PPG signal

C. Feature-based Template Generation

A typical normalized PPG signal is shown in Fig. 5. Twelve feature parameters are extracted from the PPG signals of each subject (see Fig. 5). The twelve features are defined as follows:

1. ***a*** (systolic peak)
2. ***b*** (dicrotic notch)
3. ***c*** (diastolic peak)
4. ***c/a*** (augmentation index)
5. ***(a-c)/a*** (alternative augmentation index)
6. ***d*** (the pulse width)
7. t_{pi} (pulse interval)
8. t_{pp} (peak to peak)
9. t_1 (systolic peak time)
10. t_2 (dicrotic notch time)
11. t_3 (diastolic peak time)
12. ΔT (time between systolic and diastolic peaks)

These twelve features are divided into two sets. The first set consists of first six features whereas the last six features are incorporated in the second set.

These two feature sets can be individually analyzed to develop the feature template for each subject. Variance, standard deviations etc. among the features of each set are the most popular statistical features in this kind of studies. In the present case, the enrollment template vector consisting of standard deviations of two sets of feature parameters is formulated for each subject as shown in Fig. 6, where F1 to F12 stands for feature 1 to feature 12.

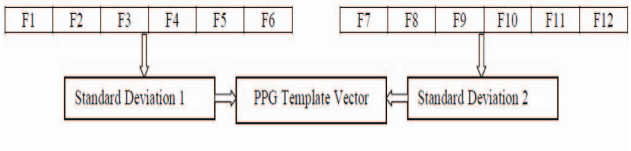

Fig. 6. PPG Template Vector

The verification is carried out by comparing the sample vector obtained at different periods of time with the template feature vector. The final decision is made based on the Linear Discriminant Classifier.

Twelve feature values and two standard deviation (Std) values of 15 subjects are shown in the Table I and Table II.

For the n classes, the intra-class matrix is given by

$$S_W = S_1 + \cdots + S_n = \sum_{i=1}^{n} \sum_{x \in C_i} \left(x - \bar{x}_i\right)\left(x - \bar{x}_i\right)' \quad (3)$$

The inter-class scatter matrix is given by

$$S_b = S_1 + \cdots + S_n = \sum_{i=1}^{n} m_i \left(\bar{x}_i - \bar{x}\right)\left(\bar{x}_i - \bar{x}\right)' \quad (4)$$

Where, m_i is the number of training samples for each class $\bar{x}_i$ is the mean for each class and $\bar{x}$ is total mean vector.

The linear transformation φ is obtained by equating generalized eigen value equation

$$S_b \varphi = \lambda S_W \varphi \quad (5)$$

TABLE I. TWELVE FEATURE VALUES OF 15 SUBJECTS

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
F1	0.889	0.807	0.785	0.854	0.908	0.923	0.848	0.943	0.896	0.913	0.934	0.901	0.877	0.883	0.801
F2	0.432	0.412	0.402	0.543	0.464	0.575	0.556	0.517	0.498	0.512	0.578	0.472	0.456	0.443	0.412
F3	0.491	0.462	0.434	0.401	0.456	0.498	0.497	0.578	0.524	0.547	0.535	0.468	0.468	0.449	0.400
F4	0.552	0.511	0.512	0.636	0.511	0.623	0.656	0.548	0.556	0.561	0.573	0.519	0.534	0.508	0.499
F5	0.448	0.489	0.488	0.364	0.489	0.377	0.344	0.452	0.444	0.439	0.427	0.481	0.466	0.492	0.501
F6	393	347	306	354	373	342	346	336	304	335	324	368	346	352	312
F7	0.014	0.010	0.012	0.011	0.014	0.013	0.011	0.012	0.012	0.014	0.014	0.014	0.012	0.013	0.012
F8	0.014	0.010	0.012	0.011	0.014	0.013	0.011	0.012	0.014	0.014	0.013	0.014	0.012	0.013	0.121
F9	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.002	0.003	0.003	0.002	0.002	0.002	0.002	0.003
F10	0.004	0.005	0.005	0.005	0.004	0.005	0.002	0.003	0.004	0.005	0.004	0.004	0.005	0.004	0.005
F11	0.004	0.012	0.010	0.008	0.006	0.006	0.005	0.004	0.008	0.007	0.005	0.007	0.007	0.007	0.008
F12	0.002	0.010	0.008	0.006	0.004	0.003	0.002	0.002	0.004	0.004	0.002	0.005	0.005	0.005	0.005

TABLE II. TWELVE FEATURE VALUES OF 15 SUBJECTS

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Subject 11	Subject 12	Subject 13	Subject 14	Subject 15
Std1	160.21	141.44	124.71	144.29	152.04	139.37	141.01	136.92	123.86	136.52	132.02	150.00	141.025	143.47	127.16
Std2	0.006	0.004	0.004	0.004	0.005	0.005	0.004	0.005	0.004	0.005	0.005	0.005	0.004	0.005	0.047

After finding the transformation φ, the classification is done in transformed space based on distance metric, such as Euclidean distance

$$d(x,y)=\sqrt{\Sigma_i(x_i-y_i)^2} \tag{6}$$

Then for a new sample Z, it is classified to

$$\arg_k = \min d(z\varphi, x_k\varphi) \tag{7}$$

Where, $\bar{x}_k$ is the centroid of the k^{th} class.

III. RESULTS AND DISCUSSION

In this work, data are collected from 15 subjects by using Biopac MP 45 data acquisition system for about three minutes. In between these three minutes of data the first one and half minutes of the data of each subject are used for training, whereas the next one and half minutes of data of each subject are used as testing.

During testing, projected new sample template vector is transformed and Euclidean distance of the transformed sample template vector from each class mean is computed.

Thus, for 15 classes 15 Euclidean distances are obtained and the smallest one among the 15 distances is used to identify the new sample vector. The proposed method is tested and validated against all 15 subjects and acquired 100% success rate. Euclidean distance for 15 sample template vector among the entire enrollment template vector is shown in Table III.

Fig. 7 shows the mean of enrollment data sets of each subject in transformed space along with the transformation axis. This figure clearly justify that transformation provide boundary for proper classification. For a large number database, where overlap occurs between classes, transformation along largest eigenvector axis proves very essential.

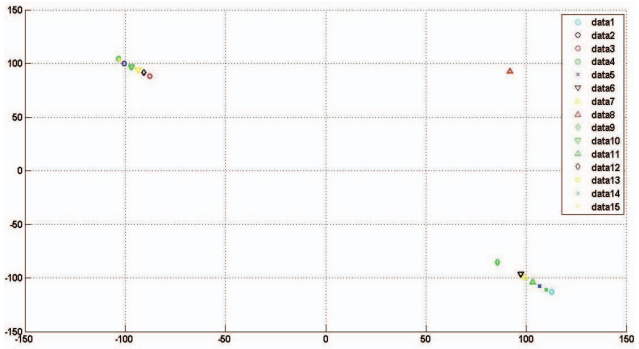

Fig. 7. Mean of enrollment data sets of each subject in transformed space along with the transformation axis

TABLE III. EUCLIDEAN DISTANCE BETWEEN SAMPLE AND ENROLLMENT TEMPLATE VECTOR OF 15 SUBJECTS

	Enrollment 1	Enrollment 2	Enrollment 3	Enrollment 4	Enrollment 5	Enrollment 6	Enrollment 7	Enrollment 8	Enrollment 9	Enrollment 10	Enrollment 11	Enrollment 12	Enrollment 13	Enrollment 14	Enrollment 15
Sample 1	2.41	301.5	284.0	306.7	8.19	22.50	18.94	206.9	38.64	296.5	12.23	287.6	291.3	2.47	304.4
Sample 2	301.9	1.42	17.85	5.05	293.7	279.4	282.9	193.0	263.2	5.41	288.6	13.30	9.58	298.3	3.80
Sample 3	284.1	17.43	5.30	22.65	275.9	261.6	265.1	179.9	245.4	12.42	270.6	4.73	8.46	280.4	21.51
Sample 4	306.0	4.64	21.93	3.76	297.8	283.5	287.0	195.2	267.3	9.56	292.9	17.54	13.81	302.6	0.93
Sample 5	8.49	293.0	275.6	298.2	3.77	14.02	10.45	200.4	30.16	288.0	5.80	281.1	284.9	4.12	297.9
Sample 6	209.6	198.5	185.8	201.3	203.8	4.93	195.6	6.70	181.8	194.9	9.67	265.9	269.6	19.38	282.7
Sample 7	17.71	283.8	266.4	289.0	9.62	4.82	1.54	193.0	20.97	278.8	6.68	268.7	272.4	16.43	285.5
Sample 8	207.2	192.7	179.8	195.6	201.1	189.6	192.7	3.12	178.5	189.0	197.5	181.5	184.3	205.2	193.9
Sample 9	38.90	262.6	245.2	267.8	30.7	16.39	19.97	177.9	5.55	257.6	25.8	249.6	253.3	35.5	266.4
Sample 10	297.5	4.14	13.45	9.19	289.3	275.0	278.6	188.9	258.9	3.45	282.9	7.59	3.88	292.6	9.31
Sample 11	13.39	288.1	270.7	293.3	5.23	9.18	5.56	197.1	25.26	283.1	3.45	275.0	278.7	10.14	291.8
Sample 12	288.5	12.98	4.49	18.23	280.3	266.0	269.6	183.2	249.9	7.97	275.5	2.86	3.66	285.2	16.65
Sample 13	292.6	8.84	8.61	14.07	284.5	270.2	273.7	186.0	254.0	3.83	279.7	4.33	3.68	289.4	12.50
Sample 14	2.94	298.7	281.3	303.9	5.41	19.80	16.19	205.6	35.88	293.7	10.25	285.6	289.3	2.63	302.4
Sample 15	305.9	4.42	21.87	3.70	297.7	283.4	287.0	195.9	267.3	9.43	292.9	17.58	13.84	302.6	3.07

A comparative study of the present work with the previously reported results is shown in Table IV.

TABLE IV. COMPARISON TABLE

Study	No of Subjects	Classification Method	Accuracy
Gu *et. al.* [12][13]	17	Euclidean Distance[12] Fuzzy[13]	94%
Kavsaoğlu *et al* [17]	40	kNN	94.4%
Jaafar *et al.* [18]	10	Bayes Network	97.5%
Proposed Work	15	LDA	100%

IV. CONCLUSION

Biometric identification system uses distinctive human characteristics like face, iris, fingerprint, gait, etc. to identify. However, all these systems may be compromised by exploitation faux credential and not proof against spoof attacks. Therefore, to stop spoofing biosignals are used as they cannot be artificially generated and guaranteed liveliness made the system advantageous over other systems. In this paper a preliminary study on Photoplethysmograpm (PPG) based biometry system is reported. Here standard deviations of temporal and special features are separately calculated for classification. The work is done under a small database of 15 persons and achieved 100% accuracy.

However, further research based on large database is needed and to present a more reliable biometric system PPG may be used as one of the attributes in a multimodal biometric system.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Trans. on Circuits Syst. Video Technol., vol.14, no.1, pp.4-20, January 2004.

[2] S. Z. Li and A. K. Jain, Handbook of Face Recognition, 2nd ed. Springer-Verlag, London, UK, August 2011.

[3] C. B. Tao and G. Liu, "An Efficient Fingerprint Preprocessing Algorithm Based on FDCT", Journal of Computational Information Systems, vol. 6, no. 12, pp. 4055 – 4063, December 2010.

[4] X. Yong, F. Lunke, D. Zhang, "Combining Left and Right Palmprint Images for More Accurate Personal Identification," IEEE Trans. on Image Processing, vol.24, no.2, pp.549-559, February 2015.

[5] L. Mirmohamadsadeghi, A. Drygajlo, "Palm vein recognition with local texture patterns," in IET Biometrics, vol.3, no.4, pp.198-206, December 2014.

[6] C. T. Chou, S. W. Shih, W. S. Chen, V.W. Cheng, D. Y. Chen, "Non-Orthogonal View Iris Recognition System," IEEE Trans. on Circuits Syst. Video Technol., vol.20, no.3, pp.417-430, March 2010.

[7] D. P. Coutinho, H. Silva, H. Gamboa, A. Fred, M. Figueiredo, "Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems," in IET Biometrics, vol.2, no.2, pp.64-75, June 2013.

[8] P. Campisi and D. La Rocca, "Brain waves for automatic biometric based user recognition," IEEE Trans. on Inf. Forensics and Security, vol. 9, no. 5, pp. 782–800, May 2014.

[9] I. Chingovska, A. Anjos, and S. Marcel, "Biometrics evaluation under spoofing attacks," IEEE Trans. Inf. Forensics and Security, vol. 9, no. 12, pp. 2264–2276, December 2014.

[10] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas., vol. 28, no. 3, pp. R1–R39, February 2007.

[11] A. Roggan, M. Friebel, K. Dorschel, A. Hahn, and G. Muller, "Optical properties of circulating human blood in the wavelength range 400-2500 nm," J. Biomedical Optics, vol. 4, pp. 36-46, January 1999.

[12] Y. Y. Gu, Y. Zhang, Y. T. Zhang, "A novel biometric approach in human verification by photoplethysmographic signals", in Proc. of the 4th IEEE Annual Int. Conf. of the Engineering in Medicine and Biology Society (EMBS), 24–26 April 2003, pp. 13–14.

[13] Y. Y. Gu, Y. T. Zhang, "Photoplethysmographic authentication through fuzzy logic", in Proc. of the IEEE EMBS Asian-Pacific Conf. on Biomedical Engineering, 20–22 October 2003, pp. 136–137.

[14] J. Yao, X. Sun, Y. Wan, "A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier", in Proc. of the 29th IEEE Annual Int. Conf.of the Engineering in Medicine and Biology Society (EMBS), 22–26 August 2007, pp. 4576–4579.

[15] P.Spachos, J. Gao, D. Hatzinakos, "Feasibility study of photoplethysmographic signals for biometric identification", in Proc. of the 17th Int. Conf. on Digital Signal Processing (DSP), 6–8 July 2011, pp. 1–5.

[16] A. Bonissi, R. D.Labati, L. Perico, R. Sassi, F. Scotti, L. Sparagino, "A preliminary study on continuous authentication methods for photoplethysmographic biometrics," in Proc. of the 2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 9-9 September 2013, pp. 28-33.

[17] A. Reşit Kavsaoğlu, Kemal Polat, M. Recep Bozkurt, "A novel feature ranking algorithm for biometric recognition with PPG signals," Computers in Biology and Medicine, vol. 49, pp. 1-14, June 2014.

[18] N. A. L. Jaafar, K. A. Sidek, S. N. A. M. Azam, "Acceleration plethysmogram based biometric identification," in Proc. of the 2015 Int. Conf. on BioSignal Analysis, Processing and Systems (ICBAPS), 26-28 May 2015, pp. 16-21.

[19] http://www.biopac.com/life-science-data-acquisition-teaching-system