Development of an Acceleration Plethysmogram based Cardioid Graph Biometric Identification

Khairul Azami Sidek, Munieroh Osman, Siti Nurfarah Ain Mohd Azam and Nur Izzati Zainal

Department of Electrical and Computer Engineering,
International Islamic University Malaysia
P. O. Box 10, Jalan Gombak, 50728 Kuala Lumpur
E-mail: azami@iium.edu.my

Abstract

The increasing identity theft cases are alarming which puts biometric as the alternative solution to combat identity crime. Recently, biosignals are proposed as biometric modalities. Thus, in this study, the development of an Acceleration Plethysmogram (APG) based Cardioid graph biometric identification is presented. A total of 10 Photoplethysmogram (PPG) data from MIMIC II Waveform Database (MIMIC2WDB) with sampling frequency of 125 Hz were obtained. The datasets are later converted to APG signal by the second order differentiation and preprocessed with Butterworth filter. Then, Cardioid based graph of APG signal was generated. Its centroid and Euclidean distance are calculated. Finally, classification is done by applying these extracted features to Multilayer Perceptron (MLP) and Naïve Bayes neural networks classifiers. Our experimentation results show that subject recognition is possible by obtaining classification accuracy of 95% for APG based Cardioid graph for both classifiers while only 85% and 70% for PPG signal in MLP and Naïve Bayes classifiers. These outcomes indicate that APG based Cardioid graph biometric identification is a feasible solution to overcome identity fraud.

Keywords: Plethysmogram, APG, Photoplethysmogram, PPG, biometric, Cardioid.

1. Introduction

Recently, identity theft or identity fraud has become one of the most popular crimes. According to the Identity Crime and Misuse in Australia Survey conducted by the Australian Institute of Criminology for the Attorney-General in May 2013, a total of 1 in 10 Australians have reported misuse of their personal information in the previous 12 months, with 1 in 5 people reporting misuse of their personal information at some time during their life [1]. The vulnerability in securing private information or identification provides enough information for fraudsters and criminals.

Moreover, a report from the CyberSecurity Malaysia showed that identity theft has increased yearly. In 2012, a total of 40% of the cybercrime related cases are from identity fraud [2]. Thus, identity theft is a serious crime which can lead a person to be at fault for something that he did not commit. Therefore, one of the strategies to combat identity crime is by using biometric recognition.

It is a general term used to describe a characteristic or a process. It automates recognition of an individual based on measurable biological (anatomical and physiological) and behavioural characteristics to identify an individual. Any measurable physiological and behavioral characteristic can be a biometric modality [3]. Examples of biometric modalities include fingerprint, face, iris and electrocardiogram (ECG).

In the past years, biosignals have been a viable identification mechanism. These signals have the capability to recognize individual via different techniques and

approaches such as the photoplethysmogram (PPG) signal. Recently, analyzing the PPG waveform has attracted interest from researcher. PPG have been widely used for biometric and medical benefits. However, based on previous literatures, PPG signal contains systolic and diastolic peaks which are hard to identify. Another biosignal called the acceleration plethysmogram (APG) is believed to provide more accurate recognition of the inflection points and an easier interpretation of the signal. APG is derived from the second derivative of PPG and acts as an indicator of the acceleration of the blood in the finger [4]. It contains lots of discriminative information which can be used for person identification. Figure 1 depicts the PPG and APG waveforms.
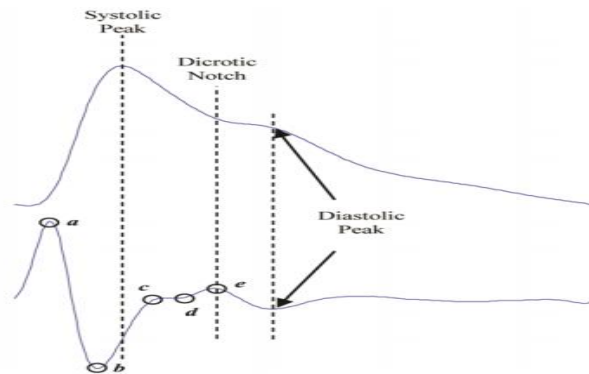

Figure 1: The upper signal is a PPG waveform consisting of one systolic and diastolic wave. The lower signal is the second derivative of a PPG. Also known as APG consisting of a, b, c, d and e waves [5].

Cardioid based graph, a feature extraction technique which have sound results in ECG analysis for subject recognition has the potential to be applied in the APG domain. The capability of applying APG based Cardioid graph in biometric identification has the potential to become a substitute or complementary entity for currently available system. Thus in this study, we will propose of applying characteristic of APG signals for biometric identification based on Cardioid graph. Based on our knowledge, little has been said pertaining to the aforementioned topic. Thus, we will propose a novel biometric identification technique in order to discriminate individuals.

2. Related Works

Recently, many researches on APG have been performed since this biosignal provide a stable baseline and can be used to separate components of the waveform clearer. Moreover, APG is believed to provide a lot of useful information [4]. Thus, in this section, we will briefly discuss on some related literature to APG signal which is divided into literatures on APG and Cardioid based graph.

Earlier, Elgendi et al in [6] have investigated if variations in APG signals can be used to measure heart rate variability (HRV) instead of the conventional method ECG signals. HRV will be calculated from APG since the peak in APG is easier to be obtained and defined resulting in more accuracy of heart rate. A numbers of PPG data is taken from 26 healthy male subjects with mean ±SD age of 27±6.9. The PPG data is measured while the subject is at rest and at sampling rate of 20 seconds. The result of this study shows that HRV indices can be calculated using APG signal as there is a strong correlation between the two HRV indices.

Elgendi et al in [7] have come out with an algorithm to detect systolic peaks under challenging conditions. The systolic-peak detection is known to be an important step in analyzing heart rate variability. A total of 30 males and 10 females of healthy respondent with ±SD age of 34.7±6.6 have participated. A PPG-equipped device is used for data

collection. Each respondent undergoes 30 minutes of triaging and resuscitating followed by 30 minutes of cooling and rest. From this study, the results showed that the proposed algorithm is robust and efficient in detecting systolic peak in PPG signal collected in challenging conditions with 99.89% of overall sensitivity and 99.84% of positive predictivity.

Later, Elgendi [8] introduced a rapid and robust algorithm to detect c, d and e waves in APG signals, specifically in a heat-stressed context. A total of 27 PPG signal was collected from a healthy male subjects with a mean ± SD age of 27 ± 6.9. The data was measured while the subject was at rest and after exercise in duration of 20 seconds for each data segment. The sampling rate while collecting data used is 200Hz. The proposed algorithms found to give a result of 99.95% sensitivity and 98.35% positive predictivity.

Next, we briefly discuss some related literature on Cardioid based graph. This method has been used by researchers as features extraction techniques in identification and medical system. Cardioid has been applied widely in identification individuals by using ECG signals.

In another paper, Sidek *et al* in [9] have proposed data mining technique on Cardioid graph based ECG biometric authentication. The ECG features are extracted by using set of Euclidean distances with the aid of data mining technique. A number of 26 ECG recordings were taken from two different public databases which comprised of 18 subjects from NSRDB and 8 subjects from MITDB. The consequence of this research shows that the classification accuracy by applying data mining technique for NSRDB is 98.60% and 98.30% for MITDB. A more efficient and rigorous method of identifying individuals for a home healthcare system is attained with the aid of Bayes Network classifier.

In another research, Sufi *et al* in [10] have performed a novel concept of Cardioid in minimizing the cardiovascular disease diagnosis and treatment time. Focusing on the increasing rate of cardiovascular related deaths, the study tried to overcome this situation by providing both faster authentication and diagnosis. A total of 30 randomly selected ECG entries from MIT BIH Database are extensively used for benchmarking algorithms pertaining to ECG diagnosis, compression and others. Two Cardioid based authentication mechanism known as Method 1 and Method 2 have been designed in this paper for authentication mechanism. From the experiment, Method 1 was found to be more accuracy with possible misclassification rate but has a higher computational expense. On the other hand, Method 2 is simpler as compared to Method 1. The proposed Method 2 requires less storage and executes faster for individual identification. The results shows that, the authentication time can be reduced from 30.64s to 0.4398s and the diagnosis time have been improved from several minutes to less than 0.5s.

However, based on these previous researches, biometric recognition using Cardioid graph on APG has never been used for individual identification. Thus, in this study, we propose of applying APG based Cardiod graph biometric identification.

3. Methodology

Recently, biometric has appeared as one of the most stable and decisive mechanism for individual identification. APG signals can be obtained by double differentiating the PPG signal. PPG signal is a non-invasive and accurate procedure to obtain physiological information. Cardioid based graph is used as extracted features. Thus, we are proposing this biological identification approaches with Cardioid based graph as a new biometric measures.

Therefore, in this section, the procedures involved in this identification system are explained and the summary of this process is illustrated in Figure 2.
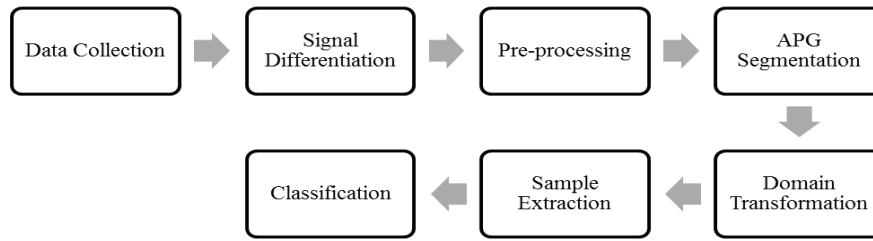
Figure 2. The Proposed Identification System

3.1. Data Collection

PPG datasets used in this study were taken from PhysioNet, a free access via the web to physiological signal archive [11]. PPG recordings duration for each subject were over ten second with sampling rate of 125 Hz. These samples were fetched from the databases of MIMIC II Waveform Version 3 Part 2 and 3 respectively.

3.2. Signal Differentiation

In order to obtain the APG signals Z[n], the PPG signals S[n] needs to be differentiated twice.

$$S'[n] = |S[n] - S[n-1]| \tag{1.1}$$

$$Z[n] = |S'[n] - S'[n-1] \tag{1.2}$$

The second derivatives of PPG or also known as APG produces peaks of a, b, c, d, and e wave which is more clearly defined compared to PPG. Thus, the interpretation process can be done easier and faster. Figure 3 shows a PPG signal converted to an APG signal.
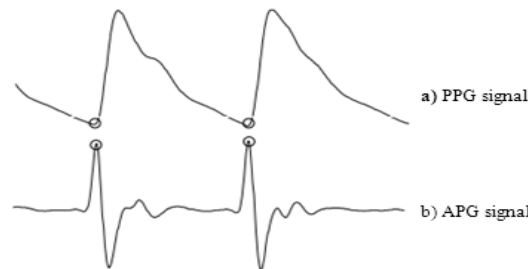

Figure 3. PPG Signal Converted to APG Signal

3.3. Pre-processing

In this part, Butterworth filter is applied to APG signal to remove high frequency noise and disturbance like baseline wander. Butterworth is used as it offers good transition band characteristics at low coefficient orders which are practical to implement. A Butterworth filter with bandpass 0.01 is implemented. The order of the filter taken is at 2nd order. Thus, a clean and more accurate APG signal is generated.

3.4. APG Segmentation

For this part, a, b, c, d, and e wave in the APG signal is defined. One complete cycle consist of a to e wave is identified. The a to e wave then is used to construct an overlap waveform of APG signals where 10 cycles of APG wave is selected.

3.5. Domain Transformation

This next step is very critical as the persistency of the classification accuracy is determined from Cardioid based person identification mechanism [9, 10]. The time series representation is lost once Cardioid is constructed as it will be replaced by points of a closed loop as shown in Figure 4.
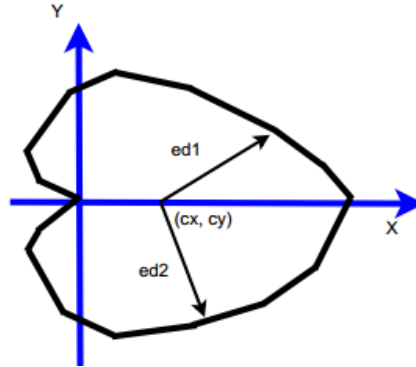

Figure 4. A Cardioid Graph

The APG signal is represented by $\mathbf{x}(t)$ as in Equation 1.3.

$$\mathbf{x}(t) = \{x(1), x(2), x(3), \dots, x(N)\} \tag{1.3}$$

where x(t) are APG signals consisting of peak waves *a, b, c, d* and *e* and *N* is the total number of APG cycles for a given time. In order to get the points for the Cardioid, the *a, b, c, d* and *e* waves are first differentiated as in Equation 1.4.

$$\mathbf{y}(t) = \mathbf{x}(t) - \mathbf{x}(t-1) \tag{3.4}$$

where $t = 1,2,3 \dots, (N-1)$ and $\mathbf{y}(t)$ is the differentiated APG dataset.

A closed loop graph is constructed based on a scattered XY graph called the Cardioid after the required vector $\mathbf{x}$ and $\mathbf{y}$ is obtained. The x-axis is the APG amplitudes of *a, b, c, d* and *e* waves and the y-axis is differentiated APG signal. Throughout this process, the time series of APG signal is changed to a two dimensional loop.

3.6. Feature Extraction

In Figure 5, centroid is the centre coordinate of the closed loop graph and the distance between the centroid to a given point on the Cardioid is known as extrema points. These two centroids and extrema points are then extracted.

$$(cx, cy) = \left[\frac{\sum_{i=1}^{N} x(i)}{N}, \frac{\sum_{i=1}^{N} y(i)}{N} \right] \tag{1.5}$$

Equation 1.5 shows how to obtain the centroid where *cx* and *cy* are the coordinate positions of the centroid in the Cardioid graph. The centroid is set as the reference point to compute the Euclidean distance, *ed(i)* as shown in Equation 1.6.

$$ed(i) = \sqrt{(cy - y(i))^2 + (cx - x(i))^2} \tag{1.6}$$

where ed(i)=ed1, ed2, ed3, …, ed(n).

The features extracted from the Cardioid based graph are then adapted to MLP classification which will be conducted in the next section.

3.7. Classification

Classification is the last step in biometric identification. This process is performed by using two commonly used classifiers which are Multilayer Perceptron and Naïve Bayes. These classification algorithms are briefly described in the next sub-sections.

3.7.1. Multilayer Perceptron

Multilayer Perceptron (MLP) method has been used for classification. MLP which consist of a few layers and a feedforward structure with an error based training mechanism is derived from the input layer, the hidden layers and the output layer. In this study, Euclidean distance of the Cardioid based graph is the input layer and the output layer will determine the class of subject. There is at least one neuron in each layer. In Figure 5, the neuron in the hidden layer is linked to both neurons in the input and output layers, where weight wi connects the input neuron xm with the hidden neuron hj, while the weight vk connects the output neuron yn with the hidden neuron hj. Input neurons xm, $1 \leq m \leq l$ is assigned to be equal to the corresponding data vector component. Next, data will propagate in a forward direction through the perceptron until the output neuron yn, $1 \leq n \leq n$ are reached.
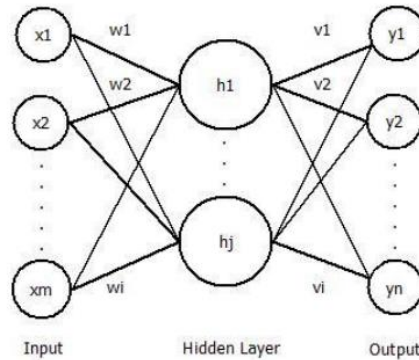

Figure 5. A Multilayer Perceptron Network

MLP neural network is chosen as the classifier because of its adaptability, noise tolerance and collective computability. Those are the main features of biometric. Therefore, the synaptic weights are determined to minimize error.

3.7. Naïve Bayes

Naïve Bayes classifier is one of the classification method used to categorize the class labels for a set of features. It is based on the Bayesian theorem and classification is performed by making a strong assumption that all the attributes are conditionally independent.
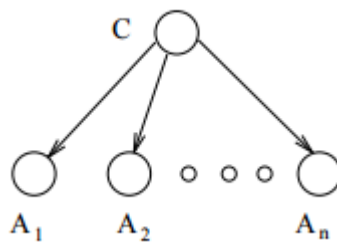

Figure 6. The structure of Naïve Bayes Network

Naïve Bayes classifier learns from training data, the conditional probability of each attribute Ai and the class label **C** as illustrated in Figure 6. Bayes rule is applied to determine the probability of the class label, **C** of the given instances $A_1, \ldots, A_n$. Then, classification is done by predicting the highest posterior probability.

Since Naïve Bayes is a conditionally independent attributes, the instances to be classified is represented by a vector $A = (A_1 \ldots A_n)$ indicating some *n* features. Using Bayes' theorem, the probability is given by:

$$P(C|A) = \frac{P(A|C)\,P(C)}{P(A)} \tag{3.7}$$

where the conditional independence assumption is as follows:

$$P(A_1, \ldots, A_n \mid C) = P(A_1|C_j) \ldots P(A_n \mid C_j) \tag{3.8}$$

The main advantage of Naïve Bayes is its simple technique and only requires small amount of training data to obtain a good estimate of the probability.

4. Experimentation and Results

A total of 10 datasets obtained from PhysioNet were tested in this study. Each dataset is taken for duration of 10 seconds with sampling rate of 125 Hz. The datasets were originally in the PPG waveform and different from each subject. Then, the signal is segmentized to 10 cycles and then differentiated twice to obtain the APG signal. We use second order Butterworth filter at frequency of 0.01 Hz to reduce noise and baseline wander of APG signal.

Figures 7, 8 and 9 depict the conversion process of PPG signal to APG signal and filtered APG signal for three subjects. The filtered segmentized APG signal consists of *a, b, c, d* and *e* waves. These waves are used for constructing the 2D Cardioid based graph.
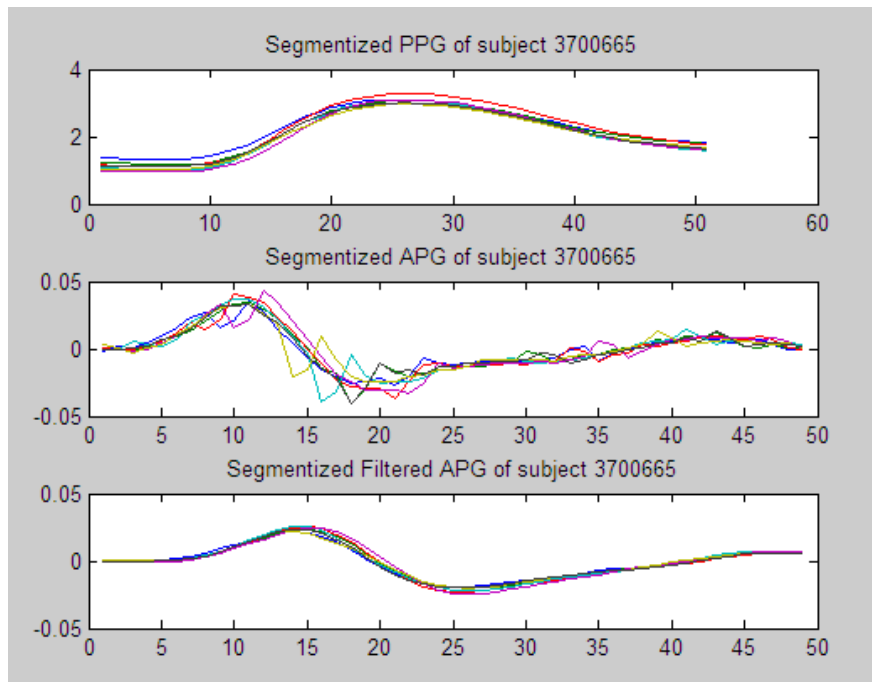

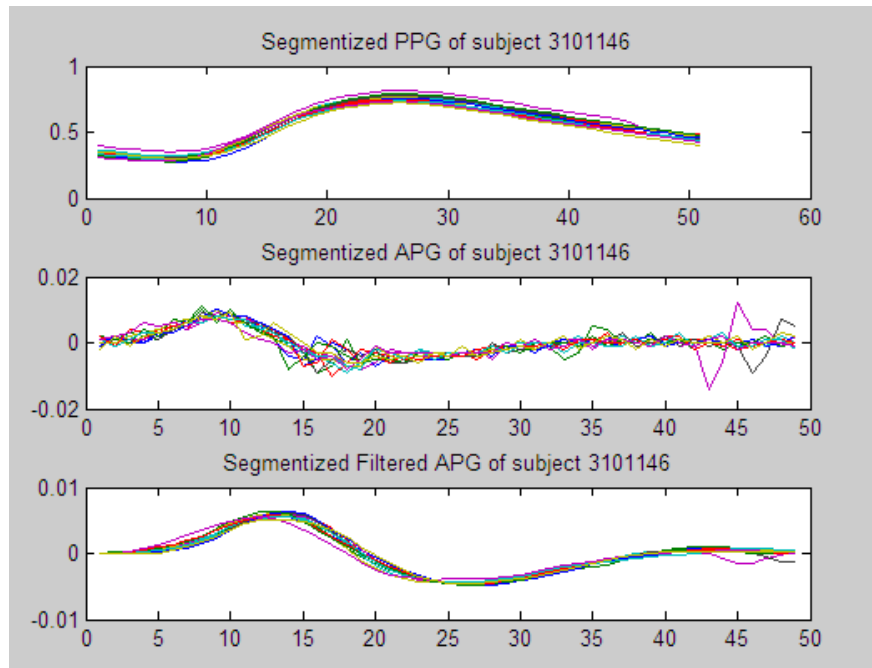
Figure 7. PPG and APG of Subject 3700665
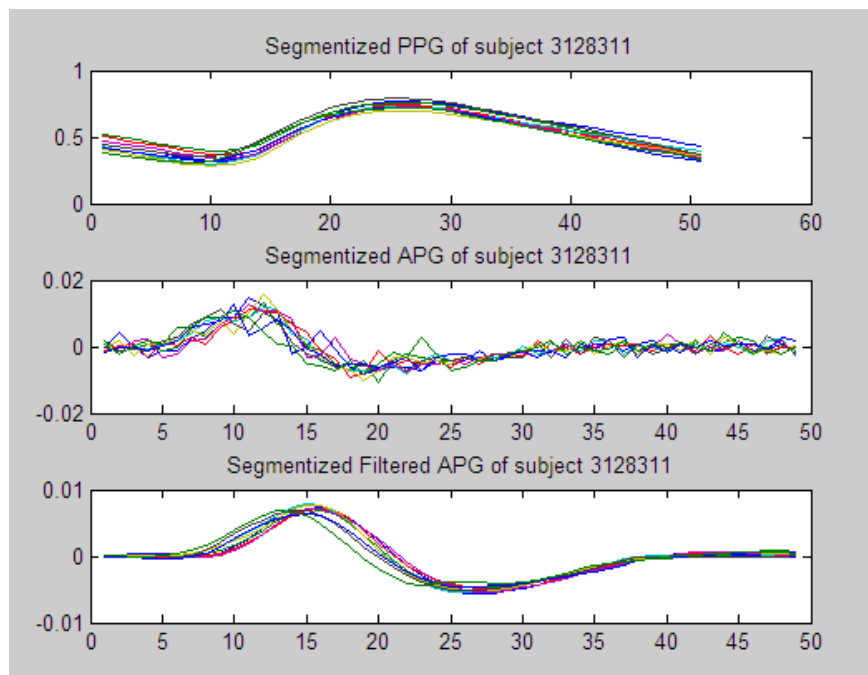
Figure 8. PPG and APG of Subject 3101146

Figure

Figure 9. PPG and APG of Subject 3128311

Figure 10, 11 and 12 illustrates the Cardioid based graph constructed from the APG signal. We observed that the Cardioid based graph for each subject is different in terms of the scale and the positions of the centroid. Generally, self-similarities are shown in the Cardioid based graph which verifies the possibility of using APG Cardioid based graph for individual identification.
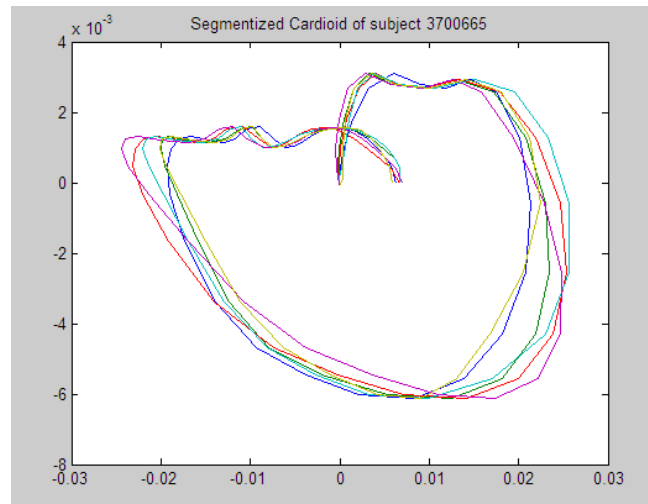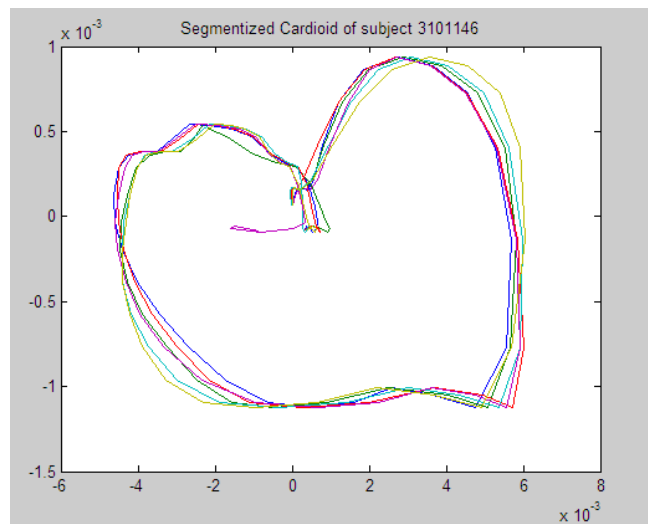
Figure 4.5. Cardioid of Subject 3700665
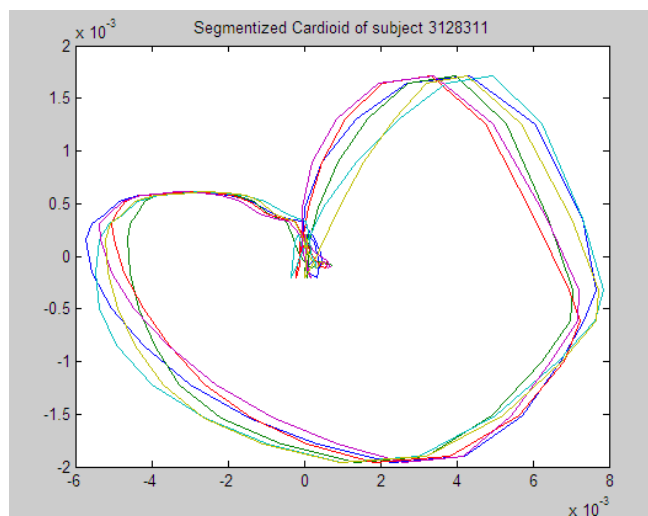

Figure 4.7. Cardioid of Subject 3101146

Figure 4.8. Cardioid of Subject 3128311

Then, we computed the centroid and the Euclidean distance of the closed loop by using Equation 1.5 and 1.6. We have selected 4 loops of Cardioid based graph for each subject where each loops contain 47 different instances. This gives a total of 188 instances for all the subjects. Half of the datasets are used as training and the remaining acts as the testing data.

Finally, MLP and Naïve Bayes neural network were applied to the instances to classify the subjects by supplying the training and the testing set. The results of the experiment are shown in Table 1.

Table 1. Classification accuracy of APG and PPG

Method	Classification Accuracy for APG	Classification Accuracy for PPG
MLP	95%	45%
Naïve Bayes	95%	55%

Based on the result, we found out that classification accuracy of APG Cardioid based graph showed significant improvement for subject recognition. All 10 subjects in APG based Cardioid graph achieved 95% average classification accuracy as compared to 85% when using PPG based Cardioid graph by applying MLP neural network classifier. Similarly, the average classification accuracy is 95% for APG based Cardioid graph and 70% for PPG signal when using Naïve Bayes classifier. This is due the facts that, APG signals contain more information as compared to PPG signals. The wave peaks a, b, c, d and e in APG signals are clearly identified in comparison to the diastolic and systolic peaks in PPG signals.

5. Conclusion

In this study, we have demonstrated an accurate and effective idea in identifying individuals by using APG based Cardioid graph with the aid of MLP and Naïve Bayes classifiers. Results of the experiment shows improvement in terms of classification accuracies of APG based Cardioid graph with a total of 95% for both classifiers as compared to 85% and 70% when using PPG based Cardioid graph. These results also suggest that the correct sample extraction technique plays an important role in determining the persistency of the classification accuracy in APG based Cardioid graph. Therefore, APG based Cardioid graph biometric identification can become an alternative solution for currently available techniques and a complementary method of recognition with the present approaches.

References

[1] R. S. Smith and A. Hutching, "Identity Crime and Misuse in Australia Survey", Australian Institute of Criminology for the Attorney-General, **(2013)**. (Accessed online on 8 October 2015).

[2] M. Z. Talib, "Identity Fraud : Platform for Cybercrime", CyberSecurity Malaysia, **(2013)**, (Accessed online on 10 October 2015) http://www.ilkap.gov.my/nlc2013/download/Nota/D2PlenarySession2.pdf.

[3] "The Facts", (n.d.). (Accessed online on 8 October 2015) ttps://stopidfraud.com.au/what-is-id-fraud/the-facts.

[4] M. Elgendi, "On the Analysis of Fingertip Photoplethysmogram Signals", Current Cardiology Reviews, **(2012)**, pp. 14-25,

[5] M. Elgendi, I. Norton, M. Brearley, D. Abbott and D. Schuurmans, "Systolic Peak Detection in Acceleration Photoplethysmograms Measured", Emergency Responders in Tropical Conditions, PLOS ONE, vol. 8, iss. 10, **(2013)**, pp. 1-11.

[6]   M. Elgendi, M. Jonkman and F. De Boer, "Applying the APG to measure Heart Rate Variability", Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference, vol. 3, **(2010)**, pp. 514-517.

[7]   M. Elgendi, M. Jonkman and F. De Boer, "Heart Rate Variability Measurement Using the Second Derivative Photoplethysmogram", Proceedings Springer-Verlag, Biomedical Engineering Systems and Technologies, **(2010)**, pp. 82-85.

[8]   M. Elgendi, "Preliminary study for localizing c, d and e waves in photoplethysmogram signals", Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE, **(2014)**, pp. 62-65.

[9]   K. Sidek, F. Sufi and I. Khalil, "Data mining technique on Cardioid graph based ECG biometric authentication", Biomedical Engineering 2011, Calgary, Canada, **(2011)**, pp. 53-57.

[10]   F. Sufi, I. Khalil and I. Habib, "Cardioids-based faster authentication and diagnosis of remote cardiovascular patients", Security and Communication Networks, vol. 4, **(2011)**, pp. 1351-1368.

[11]   M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L. Lehman, G. B. Moody, T. Heldt, T. H. Kyaw, B. E. Moody and R. G. Mark, "Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access ICU database", Critical Care Medicine, vol. 39, no. 5, **(2011)**, pp. 952-960.