Scientific
Research
Publishing

On a Feature Extraction and Classification Study for PPG Signal Analysis

Qian Wu

School of Information Engineering, Minzu University of China, Beijing, China
Email: wuqian@muc.edu.cn

Abstract

Photoplethysmography (PPG) is a low cost, non-invasive optical technology to detect the volumetric changes of blood circulation at the surface of skin. While the medical indication of components of PPG signals in the form of pulse wave are not yet fully understood, it is vastly agreed that they carry valuable pathophysiological information related to the cardiovascular system. Going beyond just dealing with frequency and time domain features of the pulse wave, as well as the first and second derivatives of the wave commonly seen in many of the relevant work, we applied a K-MEANS improved algorithm for feature extraction based on selected time domain parameters: $K1$ (systolic area), $K2$ (diastolic area) and K (entire pulse wave area). The extracted characteristic waveforms under the same light intensity could achieve average confidence level of 90% or higher. The stationary wavelet transform was adopted to further analyze the characteristic waveform by calculating the wavelet entropy; We then trained a Probability Neural Network (PNN) model using the wavelet entropy and other time domain characteristic parameters. It is found that the trained PNN model performs well in analyzing characteristic waveform to distinguish between health condition and severe arterial stenosis.

Keywords

PPG Pulse Wave, K-Means, Stationary Wavelet Transform, Wavelet Entropy, Probability Neural Network (PNN)

1. Introduction

Photoplethysmography (PPG) is an electro-optic technology to generate cardiovascular pulse wave by measuring the volumetric changes of blood circulation

at the surface of skin [1]. PPG is both clinically and individually adopted for a wide variety of application scenarios from professional diagnostics to society or home health monitoring. Numerous researches [2] [3] [4] on how to extract valuable information out of the PPG pulse wave beyond intuitive heart rate count and pulse oximetry estimation emerged recently. It is believed that the second derivative of pulse wave contains essential health-related information, hence pulse wave analysis could be of significant value in evaluating cardiovascular diseases, facilitating early detection and recognition of illnesses, and continuous health monitoring.

However due to the electro-optic nature of PPG, many factors could affect PPG signal detection [5]. For example, sensor displacement and movement due to body movement, variation of applied pressure incurred changes of magnitude of the received signal. In reality, PPG measurement usually collects excessive data to average out noises for better signal quality. Nevertheless, this inevitably could further raise difficulties for human reader of the PPG pulse wave. Peculiarity in certain pulse wave may rise simply because of affected sampling due to sensor displacement but sure causes distraction to human readers. It is therefore of practical use to extract feature waveform from vast PPG pulse wave data for the purpose of improving productivity of human readers.

We propose in the first part of this paper clustering algorithms to extract PPG pulse waves characteristics using three time domain feature parameters: $K1$, $K2$ and K, where $K1$ represents the systolic area, $K2$ denotes the diastolic area, whereas K holds the entire pulse wave area. An improved K-MEANS algorithm is adopted to extract the feature waveforms out of the pulse wave sets given the same light intensity. We present detailed algorithm implementations and the average confidence level achieved of more than 90%.

We calculate the wavelet entropy of the characteristic waveform using the stationary wavelet transform. A Probability Neural Network (PNN) model is then introduced with the wavelet entropy and six extra time domain characteristic parameters as the input for training. The trained model is tested to show the effectiveness in classification of waveforms to distinguish between health condition and severe arterial stenosis.

2. Feature Extraction

2.1. Time Domain Feature Parameters

We adopt three time domain feature parameters: $K1$, $K2$ and K, where $K1$ represents the systolic area, $K2$ denotes the diastolic area, whereas K holds the entire pulse wave area. In medical sense pulse wave area K represents characteristics of microcirculation in general but does not reflect the correlation of other feature points and areas of the whole pulse wave. We then divide the pulse wave area into 2 parts, where $K1$ is the systolic area, and $K2$ is the diastolic area.

We calculate the $K1$, $K2$ and K with reference to **Figure 1**.

K is the ratio of the area S_{ABCDE} vs. area of rectangular AHFE, denoted as below
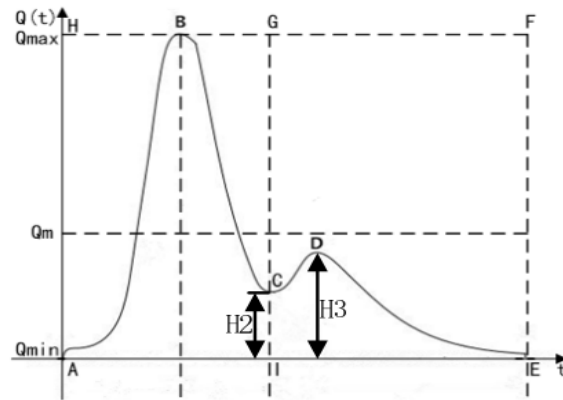
Figure 1. Pulse wave.

$$K = \frac{S_{ABCDE}}{S_{AHFE}},$$

$$S_{ABCDE} = \int_{x_A}^{x_E} G(t)dt \qquad (2.1\text{-}1)$$

where x_A denotes the start of the pulse wave segment, whereas x_E denotes the end of the segment, $G(t)$ is the function over time of the pulse wave.

Consequently, the $K1$ and $K2$ can be calculated as follows.

$$K1 = \frac{S_{ABC}}{S_{AHGI}}, K2 = \frac{S_{CDE}}{S_{IGFE}}$$

$$S_{ABC} = \int_{x_A}^{x_C} G(t)dt,$$

$$S_{CDE} = \int_{x_C}^{x_E} G(t)dt, \qquad (2.1\text{-}2)$$

$$K = K1 + K2$$

where area S_{ABC} is area starting from the X_A to the dicrotic notch, whereas S_{CDE} is the area covering dicrotic notch to X_E.
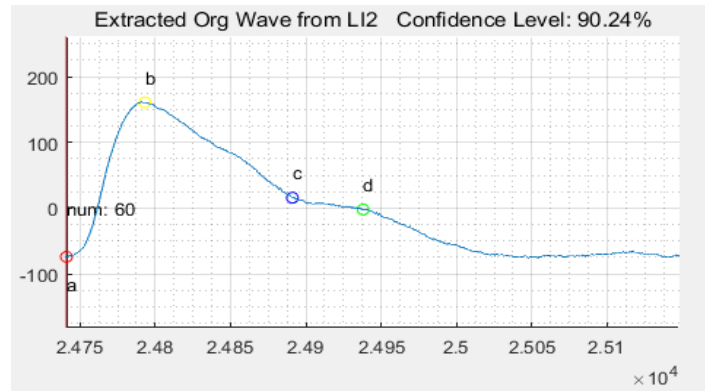
2.2. Improved K-MEANS

It is well understood that dirty data affects the clustering results with K-MEANS algorithms; while PPG measurement is prone to noise caused by many factors. We propose an improved K-MEANS algorithm by introducing updated sample center and thresholds after each round of clustering calculation in order to achieve more accurate clustering. Such an improved algorithm is less sensitive to noise and dirty data at the expense of more computing. Fortunately in our case we just have small set of data, hence it is more appropriate to land on the algorithm.

Figure 2(a) depicts the improved K-MEANS results whereas **Figure 2(b)** depicts K-MEANS results.

The confidence level of improved K-MEANS is much higher than that of the standard K-MEANS algorithms.
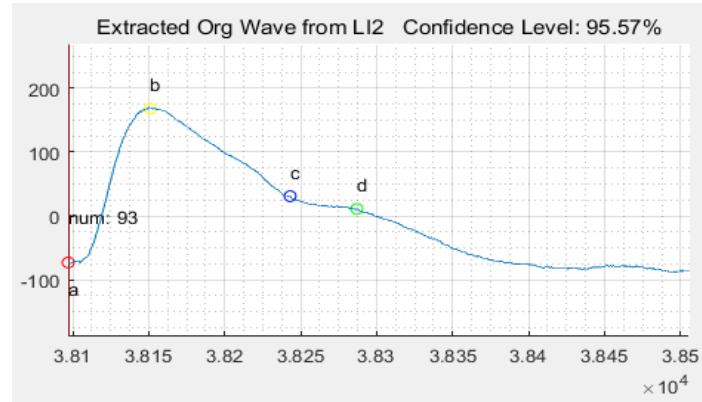
(a)

(b)

Figure 2. (a) Improved K-MEANS result; (b) K-MEANS result.

3. Stationary Wavelet Transform

Wavelet transform [6] combines both time and frequency domain together to describe the localized variation of power analysis. Wavelet provides multi-resolution analysis of pulse wave hence makes the result more insights for feature extractions. We adopted stationary wavelet transform, a.k.a., binary wavelet transform or non-decimated wavelet transform, which stops down sampling hence upon each transformation, maintains the same length as the original signal, preserve most valuable information (**Figure 3**).

3.1. Wavelet Entropy

We calculate the wavelet entropy as follows.

$$W_E = -\sum_j P_j \ln(P_j)$$

$$P_j = \frac{E_j}{E_{tot}}$$

$$E_j = \sum_k \left| C_j(k) \right|^2 \qquad (3.1\text{-}1)$$
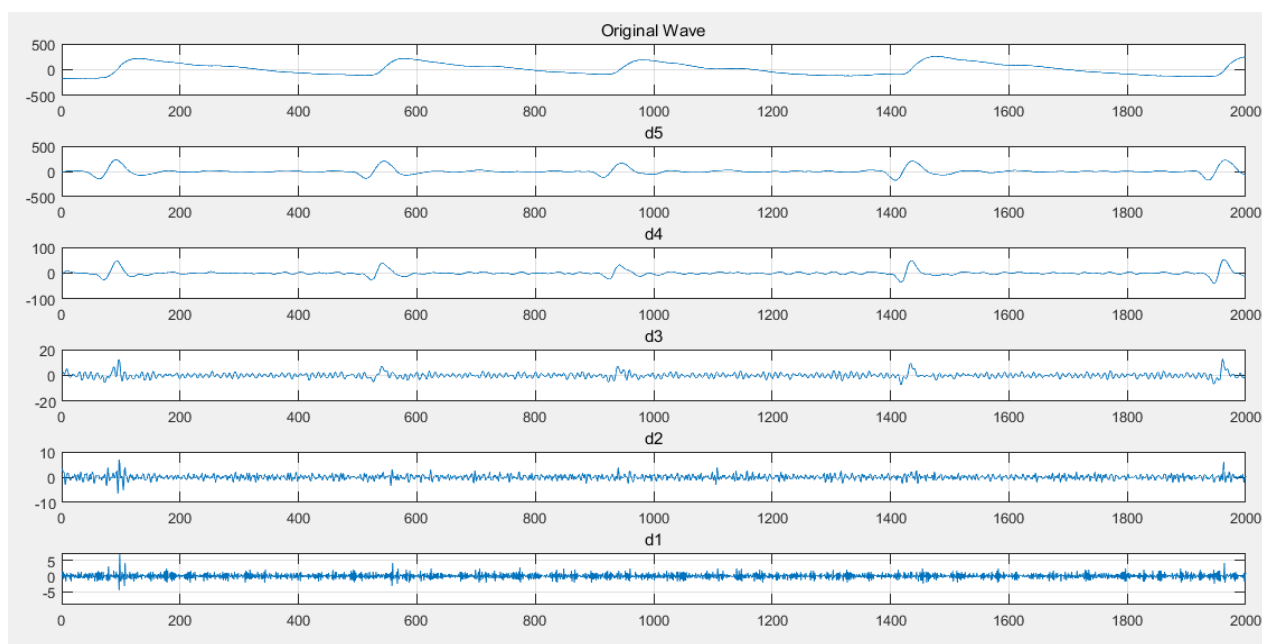
$$E_{tot} = \sum_j E_j$$

Figure 3. Original Signal vs. wavelet analysis, detailed parameters from level d1-5.

where j denotes the layers of the signal decomposition (j = 1, 2, 3, ... , 5); k is length of the original signal (k = 1, 2, 3, ..., 512); W_E denotes the wavelet entropy, E_j denotes total energy at each layer, P_j is the probability of layer j's energy vs. total energy.

3.2. Wavelet Entropy Indication of PPG Pulse Wave

- Data Preparation

PPG measurement can be affected by many factors, including pathophysiological condition and environmental condition upon test, among which age and blood pressure are key factors. We picked 23 healthy (coronary artery normal or mild stenosis) participants and 23 unhealthy (severe coronary artery stenosis) participants all at age of 50 - 70.

- Test Results

The Mean, Variance and Standard Deviation of Wavelet Entropy for healthy and unhealthy participants are listed in **Table 1**.

The fact that mean value of wavelet entropy of healthy people is less than that of unhealthy people implies that healthy people's PPG pulse wave is more stable than that of unhealthy people.

4. Classification

4.1. Probabilistic Neural Networks (PNN)

Probabilistic Neural Networks (PNN) [7] is a simple network which can be implemented using linear algebra computation and applicable to classification. As depicted in **Figure 4**, five layers are input layer, normalization layer, hidden layer, summation layer, and output layer.
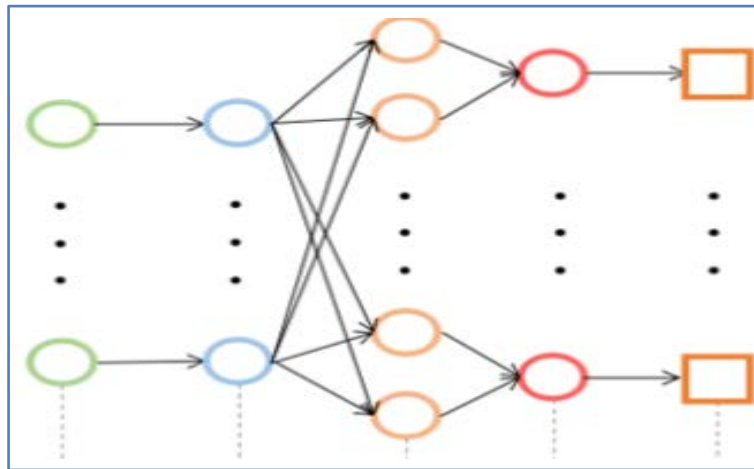
Figure 4. PNN layered structure.

Table 1. The mean, variance and standard deviation of wavelet entropy.

Wavelet Entropy	MEAN	VARIANCE	STANDARD DEVIATION
Healthy Participants	0.3639	0.0318	0.1785
Unhealthy Participants	0.4048	0.0251	0.1584

4.2. PNN Inputs

PNN Input

SIX time domain paramters: T, Pab, SI, K1, K, RI

ONE frequency domain parameters: wavelet entropy.

Where,

- T is the cycle of pulse wave; Pab = Tab/T, Tab is diff btw Xa and Xb.
- SI is height(m)/T_{bd}, T_{bd} is diff btw Xb and Xd.
- RI is H_d/H_b, H_d is vertical diff between Yd and Ya, H_b is vertical diff between Yb and Ya,
- K1 and K are defined in 2.1-2

Healthy People's Parameters are listed in **Table 2**.

Unhealthy People's Parameters are listed in **Table 3**.

It is obvious that the time domain parameters listed above for healthy and unhealthy people vary in different degree; hence it is difficult to derive any valuable information alone. As a result, we use all these time domain parameters together with wavelet entropy as inputs to the PNN for classification of PPG pulse wave.

4.3. PPG Classification w/PNN

Our test consists of 13 samples as input for training, 10 samples for classification. Results of classification are listed in **Table 4**.

It clearly demonstrated that classification results in 60% accuracy for healthy people and 80% accuracy for unhealthy ones. The reason for this is that there are clear standard to define unhealthy (stenosis) but not for the healthy ones.

Table 2. Input parameters for healthy people.

	MEAN	VARIANCE	STANDARD DEVIATION
T	0.7920	0.1192	0.0142
P_{ab}	0.2112	0.0516	0.0027
SI	7.5665	1.6493	2.7203
$K1$	0.6871	0.0349	0.0012
K	0.4015	0.0681	0.0046
RI	0.4746	0.1277	0.0163

Table 3. Input parameters for unhealthy people.

	MEAN	VARIANCE	STANDARD DEVIATION
T	0.8070	0.1382	0.0191
P_{ab}	0.2231	0.0716	0.0051
SI	6.9935	1.4744	2.1738
$K1$	0.6643	0.0897	0.0080
K	0.3947	0.0746	0.0056
RI	0.4480	0.0867	0.0075

Table 4. Classification results.

	Normal (%)	stenosis (%)	Result		Normal (%)	stenosis (%)	Result
N1	39.64	60.36	-	P1	27.38	72.62	-
N2	60.13	39.87	+	P2	30.43	69.57	-
N3	76.66	23.34	+	P3	27.83	72.17	-
N4	76.94	23.06	+	P4	35.72	64.28	-
N5	15.08	84.92	-	P5	52.20	47.80	+
N6	40.34	59.66	-	P6	21.70	78.30	-
N7	77.93	22.07	+	P7	0.60	99.40	-
N8	63.27	36.73	+	P8	91.88	8.12	+
N9	75.76	24.24	+	P9	46.71	53.29	-
N10	39.31	60.69	-	P10	42.67	57.33	-

Where, N denotes healthy people, P denotes unhealthy people; "+" denotes coronary artery normal; "-", denotes severe coronary arterystenosis; "_" denotes misclassification.

5. Conclusion

The feature extraction and classification methodology for PPG signals using improved K-MEANS improved algorithm, stationary wavelet transform and PNN modelling is easy to implement and effective to use. Time domain parameters and frequency domain wavelet entropy are appropriate data set for PNN modelling to achieve acceptable classification results. We see all this as a start for further work to gain more insights into pathophysiological indication of PPG pulse wave.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C. and Nazeran, H. (2018) A Review on Wearable Photoplethysmography Sensors and Their Potential Future Applications in Health Care. *Int J Biosens Bioelectron*, **4**, 195-202. https://doi.org/10.1109/IEMBS.2006.4398399

[2] Cohn, J.N., Finkelstein, S.M., McVeigh, G.E., *et al.* (1995) Noninvasive Pulse Wave Analysis for the Early Detection of Vascular Disease. *Hypertension*, **26**, 503-508. https://doi.org/10.1161/01.HYP.26.3.503

[3] O'Rourke, M., Pauca, A. and Jiang, X.-J. (2001) Pulse Wave Analysis. *Br J Clin Pharmacol.*, **51**, 507-522. https://doi.org/10.15406/ijbsbe.2018.04.00125

[4] Zhang, G., Kong, X. and Liao, S. (2008) Pulse Wave Analysis for Cardiovascular Information Monitoring in Patients with Chronic Heart Failure: Effects of COQ10 Treatment. *Montreal*: Bio-Engineering 2008. https://doi.org/10.1016/B978-0-12-816514-0.00014-X

[5] Bolanos, M., Nazeran, H., Haltiwanger, E., *et al.* (2006) Comparison of Heart Rate Variability Signal Features Derived from Electrocardiography and Photoplethysmography in Healthy Individuals. *Engineering in Medicine and Biology Society*, **1**, 4289-4294. https://doi.org/10.1046/j.0306-5251.2001.01400.x

[6] Weng, H. and Lau, K.-M. (1994) Wavelets, Period Doubling, and Time-Frequency Localization with Application to Organization of Convection over the Tropical Western. Pacific. *J. Atmos. Sci.*, **51**, 2523-2541. https://doi.org/10.1175/1520-0469(1994)051<2523:WPDATL>2.0.CO;2

[7] Mohebali, B., Tahmassebi, A., Meyer-Baese, A. and Gandomi, A.H. (2020) Probabilistic Neural Networks: A Brief Overview of Theory, Implementation, and Application. Elsevier, 347-367.