Preprocessing PPG and ECG Signals to Estimate Blood Pressure Based on Noninvasive Wearable Device

Mengting Wang, Chao Huang, Hang Chen & Shuming Ye
College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China
Key lab of Biomedical Engineering of Ministry of Education, Zhejiang University, Hangzhou, Zhejiang, China
Zhejiang Provincial Key Laboratory of Cardio-Cerebral Vascular Detection Technology and Medicinal Effectiveness Appraisal, Hangzhou, Zhejiang, China

ABSTRACT: Accurate systolic and diastolic blood pressure measurement is still an open problem in biomedical engineering. This paper introduces a method how to preprocess the PPG and ECG signals to get a series of parameters to estimate blood pressure. It developed a threshold detector to extract characteristic points of PPG and ECG waves and calculated several useful features such as pulse wave velocity. The paper also introduced a BP-feature based model by using those features to estimate SBP and DBP. The accuracy of detection algorithm reached 98.2% by comparing with SFM database. The estimation results were validated by a large-scale dataset we acquired by strict experiment procedures. The average deviation error in estimating SBP and DBP was 0.25 and 2.24 mmHg respectively. The standard deviation between the measured and predicted blood pressure was 8.92 mmHg for systolic pressure and 8.13 mmHg for diastolic pressure, which met the standard of AAMI. The results indicated that the BP-feature based model has a reliable estimation of blood pressure.

Keywords: feature extraction; SBP; DBP; BP-feature based model

1 INTRODUCTION
The discussion of cardiovascular diseases and hypertension is popular in recent decades in both developing and developed countries. The increased stiffness of arterial wall will rise the risk of cardiovascular morbidity without doubt[1]. Consequently, it is quite urgency to monitor noninvasive and continuous blood pressure to improve the management and assessment of cardiovascular and hypertension diseases.

It has been years for researchers in the worldwide studying the measurement of noninvasive and continuous blood pressure. Various kinds of commercial devices which can monitor blood pressure can be seen on the market. Most of the existing devices require to get radial arterial pressure and figure pulse wave respectively based on the technology of vascular unloading method[2]. Some of them, such as T-line (Ten-sys Medical System[3]) and CNAPTM monitor 500 (CNSystems Medizintechnik AG[4]), satisfy the accuracy standard of the Association for the Advancement of Medical Instrumentation (AAMI[5]). Nevertheless, these devices are found to be difficult to be widely used in clinic because of not convenient to use, high costs to be manufactured and complicated mechanical structure.

In order not to use those complex and expensive sensors, we make use of Photoplethysmography (PPG) for blood pressure measurement. The PPG waveform will be changed with the volume changes of the blood vessel, thus PPG can indicate relevant information of peripheral blood circulation. Through analyzing PPG waveform we can get the information of pulse wave velocity (PWV) or pulse transmit time (PTT) which has been used in studying for years since the mature technology. In our research, the figure pulse waves are collected using our noninvasive and wearable health monitoring system based on PPG.

PPG and ECG signals are often influenced by movement of human body unavoidably, which may cause false alarms. Researchers have found all kinds of strategies to reduce these kinds of alarms, using median filtering, multi-parametric analysis, machine learning and signal quality assessment technology[6].
In this paper, a reliable method is presented to pre-process both PPG and ECG signals to extract characteristic points. Based on these points we calculated morphology parameters of PPG waveforms to estimate SBP and DBP based on BP-feature model with the knowledge the human cardiovascular model. Although this new method is still unreliable to be applied on medical diagnoses, the results meet the standard of AAMI. Consequently, it is worth studying in the direction of the fields of noninvasive and continuous blood pressure monitoring.

2 MATERIALS AND METHODS

The proposed noninvasive blood pressure estimate method involves six parts: (1) data collection; (2) signal filtering; (3) feature extraction; (4) Pulse wave velocity calculation; (5) Morphology of PPG analysis; (6) BP-feature based model.

2.1 Data collection

The database consists of our previous experiments which have been conducted on 1015 individuals of a population-based sample, including 446 males (mean age at 40.5, range 18-83) and 569 females (mean age at 42, range 15-80). Before we begin the measurement, we make sure that no one in our experiments has cardiovascular related diseases. All the participants were instructed about the steps of our experiments and gave informed consent to participate in before. Briefly, the procedures of our experiments were arranged into four steps.

Step 1: the information of subjects including sex, age, height, weight were registered on the books and then all the subjects need to sit in peace for five minutes.

Step 2: use our device which was studied previously by the students in Key lab of Biomedical Engineering of Ministry of Education, Zhejiang University to measure blood pressure of the subjects. During the measurement, talking is forbidden to the test subjects. The results are regarded as the standard blood pressure of subjects.

Step 3: wear our noninvasive wearable device on the figures and the body of the subjects. We collect both PPG and ECG signals for one minute at the sample of 500hz. The waves we collected are used to build our BP-feature based model.

Step 4: repeat step 2 again to assure that the blood pressure of the subjects is in a stable station, otherwise we repeat step 2 again so that we can make sure we measure the accurate blood pressure.

2.2 Signal filtering

It is known that during the measurement inevitably there are kinds of noises like baseline drift and 50 HZ power frequency interference. Motion artifact is also a kind of usual occurrence accompany measurement. The collected PPG and ECG signals were analyzed in MATLAB by a tailor-made program to get the clear waveform.

2.2.1 Basic noise filtering

We designed a digital butter-worth filter with pass frequency 0.5Hz and stop frequency 20 Hz to ECG signals to filter the respiration interference and electrical noise. Median filter was applied to remove baseline shift. PPG signals were filtered through a 3-order low-pass Bessel filter. Notch filter was used to eliminate 50 Hz power frequency noise.

2.2.2 Motion artifact recognition

According to our experiments, we found that PPG signals were very sensitive to motion artifact such as sudden movement or spike of circuit voltage. However, it is quite difficult to remove motion artifact by a normal filter because motion artifact don’t have a predefined narrow frequency band and the spectrum often overlaps that of the desired signals.

In this study, we recognized movement artifact based on the difference of heart rates calculated by PPG and ECG signals respectively. Because PPG and ECG signals were collected synchronously, we first divided both signals into segments of 10 seconds. Then the peaks of PPG and ECG signals were detected to compute heart rates respectively. Since the ECG signals were less polluted by movement artifact, heart rate calculated by the peaks of ECG signals was regarded as the real heart rate. It was compared with the result calculated by peaks of PPG signals to distinguish whether the PPG was influenced by motion artifact. The criteria to judge were based on a threshold of 3.6 bpm by a ROC analysis. Once we assure that the segment was corrupted by the motion artifact, we abandon this segment of signals of both PPG and ECG. Only in this way can we get the accurate estimation of heart rates and pulse wave velocity.

2.3 Feature extraction

2.3.1 Extracting onsets, systolic peaks and dicrotic notches of PPG signals

It is much difficult to extract characteristic points than beat computation. Using systolic peaks to compute pulse beats is easy and acceptable. Nevertheless, because onsets and ends of pulse beats have quite weak amplitudes and are more susceptible to noises and movement artifact, it is more challenging to extract full characterization of PPG signals than just finding systolic peaks.

This paper implemented an algorithm on MATLAB to check out the onsets, systolic peaks and dicrotic notches of PPG signals. This threshold detector of the feature points was on the basis of a series of nonlinear...
transformation and first derivative operator\(^9\). Besides, several decision logics were used to handle pathophysiological complexity and instrumental unreliability in large scale PPG signals.

In fact, onset of every pulse beat means the time that blood begins to eject from heart to aorta and diastolic notch means the closure of aortic valve. So we can get a conclusion that through the positions of these points we can get knowledge of cardiac function and vascular condition. Figure 1 fully describes the procedures of detecting PPG characteristic points.

When we got the clean PPG waveforms, amplitude and interval thresholds were estimated adaptively. Firstly, PPG signals were segmented into equal pieces and a selective window whose length was 2.5 seconds was applied to the start of every segment. Within the selected window, amplitude and interval thresholds were estimated and averaged as initial thresholds.

![Figure 1. The procedures of detecting characteristic points of PPG waves.](image1)

Meanwhile, when we finished the derivative calculation, we sought zero-crossing points and pairs of inflection. Figure 2 shows one typical PPG waveform and its derivatives. In the first derivation, the onsets were relevant with zero-crossing points before a maximal inflection while systolic peaks were after the reflection. Amplitude of PPG signals and the interval norm were used to pick out all the possible onset points and candidate systolic peaks. If those points were qualified, steps were taken to find diastolic notches. If not qualified, the program would step backward to modify thresholds of both amplitude and interval.

There was a beat evaluation to find out a new beat. During characteristic detection, we found different people have very different positions of diastolic notches. Once a new beat was found, the inflection detection of candidate diastolic notch was applied\(^10\). It was assumed that there was a diastolic notch after every systolic peak. A searching window whose length was more than 40ms and less than 400ms was applied to the pairs of inflection and zero crossing points. Diastolic notch was defined as the first zero crossing position after the secondary inflection in every beat. In practice, we found sometimes diastolic notches not existed in some pulse beats.

![Figure 2. PPG waveform and its derivatives.](image2)

![Figure 3. Characteristic points detected of a typical PPG waveform.](image3)

The accuracy of our algorithm to detect characteristic points of PPG was validated and evaluated by an open database namely SFM database\(^10\) which was built up by using data in Fantasia and SLP databases\(^11\) to evaluate beat detector. There were 36 pieces of PPG waveforms in database SFM. All onsets, systolic points and diastolic notches of PPG waveforms have been annotated by well-trained engineers and approved by medical experts. We compared the number
of onsets, systolic peaks and dicrotic notches we found by our algorithm and annotations approved in the database, and results were listed in Table 1. As we can see in the Table 1, our algorithm got reached to an average distinguished error rate at 1.8%. As a matter of fact, taking amplitude and interval criteria into account made the detection of characteristic points more reliable and accurate.

<table>
<thead>
<tr>
<th></th>
<th>Annotations</th>
<th>Algorithm Found</th>
<th>Error(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onsets</td>
<td>2564</td>
<td>2569</td>
<td>0.69%</td>
</tr>
<tr>
<td>Systolic peaks</td>
<td>2564</td>
<td>2506</td>
<td>1.10%</td>
</tr>
<tr>
<td>Dicrotic notches</td>
<td>2564</td>
<td>2409</td>
<td>3.70%</td>
</tr>
</tbody>
</table>

### 2.3.2 Extract R-wave of ECG signals

With the clean ECG waveform we processed in section 2.2, R-wave detection algorithm was implemented. Firstly, the ECG data of each person was divided into the length of the period of 2.5 seconds. Secondly, each piece was studied with technical standard according to the length of interval of R-R peaks, amplitude of peaks and the morphology of R-wave. After we found the periodically repeated similar ECG signals, starting points of each ECG segment were changed according to the quality R-wave checking procedures. After we got a series of candidate detected peaks, a strict criterion was applied to select real R-peaks. Then we got a new and correct sample of ECG segment. By using those characteristics of selected peaks, modification was taken to change the starting or end of the next segment of ECG data[12]. In the end, falsely detected R-peaks were denied and we got real R-peaks of the whole ECG signals. By using the average R-R interval, we can easily calculate the heart rate, which is a common index of today’s wearable health monitors. Figure 4 illustrates the results of R-wave detection.

![Figure 4. Detection of R-wave.](image)

### 2.4 Pulse wave velocity calculation

Pulse wave velocity (PWV) is an important indicator showing the stiffness of blood vessel. With the noninvasive wearable device we used during the experiments, PPG and ECG signals were acquired synchronously. To calculate pulse wave velocity, pulse transit time was got first. Figure 5 illustrates pulse transit time (PTT) which is defined as the time difference between the systolic peaks of PPG and R peaks of ECG waveforms. PTT was computed by every pair of peaks of PPG and ECG and the average value was regarded as one person’s typical PTT. Beat-to-beat pulse transmit time was calculated as Equation 1:

\[
PWV = \frac{\text{Distance}}{\text{PTT}}
\]

By statistics, the distance between the heart to the figure is approximately one fifth of height[3].

![Figure 5. PTT on PPG and ECG waveform.](image)

### 2.5 Morphology of PPG analysis

In fact, it is inaccurate and incorrect to use PWV as a unique parameter to estimate blood pressure. Frederick Akbar Oratio Mahomed in 1872 said since the information which pulse affords is so important, surely it must be our advantage to appreciate fully all it tells us. In this paper, we analyzed the morphology parameters of PGG in seven indexes: $T_1$ (up-time), $T_2$ (down-time), $S_1$ (up-area), $S_2$ (down-area), $S_3$ (all-area). These features are illustrated in the following Figure 6. $T_1$ is the average time between the point of aortic valve openings with the blood of the left ventricle being discharged and the appearance of percussion wave. $T_2$ is the mean time between percussion wave and the end of the PPG wave. $S_1$ (up-area) is defined as the area of PPG wave during $T_1$, and $S_2$ (down-area) is regarded as the area of the PPG wave during $T_2$. $S_3$ (all-area) is defined as the area of PPG wave in one period. We also calculated RRI which is the average R-R interval of PPG signals and HR, which is short for heart rate. In addition, we also took parameters such as gender, weight, height, BMI (Body Mass Index) into account when blood pressure was estimated.

![Figure 6. Morphological parameters of PPG ($S_1$, $S_2$, $T_1$, $T_2$).](image)
2.6 BP-feature based model

2.6.1 KNN classifier

To estimate SBP and DBP, K-nearest-neighbor (KNN) classification algorithm was adopted to cluster the data into two groups (normal blood pressure and high blood pressure). KNN classification is one of the most fundamental and simple classification methods\cite{13}.

In this paper, KNN classifier algorithm was carried out to divide people into groups of high blood pressure (SBP>140mmHg) and groups of normal blood pressure (SBP<140mmHg and 90mmHg). In the first place, feature selection is a key factor in this algorithm because too much redundancy characteristics will increase the complexity of computation. Among all the features we have got, statistical analysis was used to facilitate feature selection. We made use of Pearson correlation coefficients to find out the linear relationship between features and blood pressure. All Pearson coefficients are listed in the following Table 2.

The result shows human blood pressure is associated with gender, age, weight, PWV, up-area (S\textsubscript{1} in section 2.5) and BMI. It is obviously that height, RRI, HR, T\textsubscript{1}, S\textsubscript{2} and S\textsubscript{3} fail to be significant in blood pressure estimation. And with the help of these relevant features, a classifier, K-nearest-neighbor (KNN) classification was adopted to predict whether blood pressure was normal or in a high level. A method of 10-fold cross-validation was used to evaluate how our algorithm worked. The result shows the accuracy rate of the classifier based on k-nearest-neighbor reaches 87.4%.

2.6.2 Data fitting

After divided all the training data into two groups, linear regression was applied to fit systolic blood pressure and diastolic blood pressure respectively of both two groups. Regression analysis is a statistical process for estimating the relationships among variables. A software named WEKA which was a collection of machine learning algorithms for data mining tasks was used to get our final equations to compute blood pressure.

3 RESULTS AND EVALUATION

3.1 Fitting results to estimate blood pressure

For people who were divided to the group of normal blood pressure according to their features, SBP was calculated as following equation (2) and DBP was computed as equation(3):

\[
\text{SBP} = -0.286 \times \text{gender} + 0.17 \times \text{age} + 0.2 \times \text{weight} + 0.21 \times \text{bmi} + 1.88 \times \text{pwv} - 77.63 \times \text{s1} - 36.94 \times \text{s2} + 29.15 \times \text{s3} + 138.84
\]  

\[
\text{DBP} = -3.82 \times \text{gender} + 0.12 \times \text{weight} + 0.13 \times \text{bmi} - 0.03 \times \text{rri} + 3.42 \times \text{pwv} - 64.93 \times \text{s1} - 59.83 \times \text{s2} + 48.51 + 106.91
\]  

For people who were divided into the group of high blood pressure, SBP and DBP were calculated as equation (4) and (5) respectively:

\[
\text{SBP} = 3.67 \times \text{gender} - 0.23 \times \text{weight} + 0.95 \times \text{bmi} + 0.17 \times \text{ptt} + 9.67 \times \text{pwv} - 83.4 \times \text{s1} + 116.86
\]  

\[
\text{DBP} = -4.1 \times \text{gender} - 0.37 \times \text{hr} + 0.09 \times \text{t1} + 146.9687
\]

Note that, the gender we used in equations means 1 for women and 0 for men. RRI, PWV, S\textsubscript{1}, S\textsubscript{2}, S\textsubscript{3}, HR were described in section 2.5.

3.2 Evaluation of the BP-feature model

There are two indexes to assess the accuracy of our BP-feature model of estimating systolic blood pressure and diastolic blood pressure. One index is average deviation, which means the average value of the differences between the predicted blood pressure and the measured blood pressure. The other index is standard deviation, which is usually a measure of a set of data dispersion degree of the average. Standard deviation is computed as Equation 7. N is the number of data, \(x_i\) is difference the predicted blood pressure and the measured blood pressure.

\[
\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}
\]  

After applying the equations we got in section 3.1 to nearly 512 pieces of testing data, the results of our method to estimate the systolic and diastolic blood pressure is as described in Table 3.

<table>
<thead>
<tr>
<th>Blood pressure</th>
<th>Average deviation</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP</td>
<td>0.25</td>
<td>8.92</td>
</tr>
<tr>
<td>DBP</td>
<td>2.24</td>
<td>8.13</td>
</tr>
</tbody>
</table>

Table 2. Pearson correlation coefficients between features and blood pressure.

<table>
<thead>
<tr>
<th></th>
<th>Sex</th>
<th>Age</th>
<th>Height</th>
<th>Weight</th>
<th>BMI</th>
<th>HR</th>
<th>PWV</th>
<th>T1</th>
<th>S1</th>
<th>S2</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP</td>
<td>-0.104</td>
<td>0.507</td>
<td>-0.08</td>
<td>0.258</td>
<td>0.005</td>
<td>0.008</td>
<td>-0.22</td>
<td>0.15</td>
<td>-0.357</td>
<td>-0.065</td>
<td>-0.115</td>
</tr>
<tr>
<td>DBP</td>
<td>-0.19</td>
<td>0.358</td>
<td>0.001</td>
<td>0.282</td>
<td>-0.17</td>
<td>0.166</td>
<td>-0.263</td>
<td>0.061</td>
<td>-0.225</td>
<td>0.0278</td>
<td>-0.223</td>
</tr>
</tbody>
</table>
The mean deviation (0.25 mmHg for SBP and 2.24 mmHg for DBP) and standard deviation (8.92 mmHg for SBP and 8.13 mmHg for DBP) satisfied the criteria of AAMI. The results mean that our method of computing the characteristics of PPG and ECG to estimate normal and high SBP and DBP for people with no evidence of cardiovascular disease is valid. In addition, our BP-feature based model can estimate SBP exceeding 150 mmHg.

In general, the errors of estimation may come from inaccurate computation of PWV, other parameters such as S1, and influences of physiological and environment. Nevertheless, the results in Table 3 show the errors from various sources mentioned above are within acceptable range.

4 DISCUSSION

4.1 Using photoplethysmography sensors to get noninvasive signals

PPG waveforms can show a lot of pathophysiological information of cardiovascular circulation system. Accurate measurement of PPG in daily life is quite necessary and is not a new idea in health care monitoring system. However, most products use inflatable cuff measurement methods which can bring uncomfortable feelings during daily measurement, while we make use of photoplethysmography sensors in order to get PPG signals in a noninvasive way. It is easy to use and can be operated at any time even during sleep in a safety way which can help us to realize the goal of monitoring blood pressure in 24 hours.

4.2 Developing a BP-feature based model

The BP-feature based model is the first time to use various features like PWV, up-area (S1) to estimate blood pressure in normal and high conditions. We calculated the correlations of all the features with blood pressure and picked out factors may have influence on blood pressure. In addition, the BP-feature based model proposed in this paper has been proved to be accurate and reliable to monitor human health according to the results described previously.

Algorithms reported in previous papers were usually on the basis of very small datasets. Sometimes the accuracy of the algorithm was unable to reach the international criteria. As a matter of fact, because of individual variation, pathophysiological complexity, and instrumental errors, the reliability of those algorithms may face big challenges when been applied to the large-scale datasets. However, the database we used to build and evaluate the BP-feature based model is consist of systolic and diastolic blood pressure values in a large scale which were acquired by strict experiment procedures.

4.3 Limitations and future work

Because of the constraint of time, personnel and the environment of experiment, we just conducted our experiment on the white race to validate our estimate methodology. The behavior of our BP-feature based model on the white race and the black race is still unknown. In the future research, we hope to conduct more experiment on other races whose ages are very different.

Moreover, dicrotic notches contain a lot of information of vascular condition. We have already extract dicrotic notches of PPG waveforms but we didn’t make full use of them. In the future work, we can calculate more morphology parameters of pulse wave signals to make more accurate estimation.

5 CONCLUSION

In this paper, pulse wave and ECG wave are acquired in a noninvasive way according to different ages and genders. And then after processing the signals and applying algorithms on processed data to extract features of biological waveform, a BP-feature based model was built to estimate blood pressure using noninvasive wearable device. This method is still being studied and will be improved to provide people a more accurate blood pressure estimation which has a large application in many aspects. A smart watch with a PPG sensor is commercially available for monitoring heart rate and blood pressure.

REFERENCES

[3] Yan Chen, Changyun Wen, Guocai Tao, and Min Bi 2012. Continuous and noninvasive measurement of systolic and diastolic blood pressure by one mathematical model with the same model parameters and two separate pulse wave velocities. Annals of Biomedical Engineering, 40(4).


