

Remote Estimation of Continuous Blood Pressure by a Convolutional Neural Network Trained on Spatial Patterns of Facial Pulse Waves

Kaito Iuchi

Graduate school of science and engineering,
Department of imaging science,
Chiba University,
JAPAN
lvktcd@gmail.com

George C. Cardoso
FFCLRP,
Physics Department,
University of São Paulo,
BRAZIL
gcc@usp.br

Norimichi Tsumura

Graduate school of science and engineering,
Department of imaging science,
Chiba University,
JAPAN
tsumura@faculty.chiba-u.jp

Ryogo Miyazaki

Graduate school of science and engineering,
Department of imaging science,
Chiba University,
JAPAN
wannkombrr@chiba-u.jp

Keiko Ogawa-Ochiai

Department of General Medicine,
Hiroshima University Hospital,
JAPAN
ikkandoo@gmail.com

Abstract

We propose a remote method to estimate continuous blood pressure based on spatial information of a pulse wave at a single point in time. By setting regions of interest to cover a face in a mutually exclusive and collectively exhaustive manner, RGB facial video is converted into a spatial pulse wave signal. The spatial pulse wave signal is converted into spatial signals of contours of each segmented pulse beat and relationships of each segmented pulse beat. The spatial signal is represented as a time-continuous value based on a representation of a pulse contour in a time axis and a phase axis and an interpolation along with the time axis. A relationship between the spatial signals and blood pressure is modeled by a convolutional neural network. A dataset was built to demonstrate the effectiveness of the proposed method. The dataset consists of continuous blood pressure and facial RGB videos of ten healthy volunteers. A comparison of conventional methods with the proposed method shows superior error for the latter. The results show an adequate estimation of the performance of the proposed method, when compared to the ground truth in mean blood pressure, in both the correlation coefficient (0.85) and mean absolute error (5.4 mmHg).

1. Introduction

Blood pressure is an important biomarker that reflects health status. Health risks due to hypertension include heart and kidney failure, and health risks due to hypotension include a decline in metabolism and brain function. It is essential to measure blood pressure continuously for health management. Moreover, an instantaneous fluctuation of blood pressure contains critical information. For example, the reserve capacity can be observed by the resilience of blood pressure to a postural change. Thus, health care requires high temporal resolution in blood pressure measurement. Continuous blood pressure measurement requires the use of protrusive and expensive equipment, except for a few new ideas still under investigation, such as the use of finger oximeters for blood pressure evaluation [1,2]. It is desirable to read blood pressure using simple and non-contact equipment. Recently, researchers have been intensively studying remote methods to estimate blood pressure using RGB cameras. Jeong et al., Fan et al., and Huang et al. [3,4,5] focused on a correlation relationship between pulse transit time (PTT) measured remotely and blood pressure based on the Moens-Korteweg equation [6]. There are two main limitations to these methods: a requirement for simultaneous capturing of a face and a

palm with an RGB camera, and a requirement for a very high temporal resolution of an RGB camera. Sugita et al. and Buxi et al. focused on the relationship between characteristics of pulse wave contours and blood pressure measured remotely [7,8]. These methods provide relaxation of the two requirements in PTT-based methods. First, the methods only require the capturing of a single part, which is a face or a palm. Second, the methods require the relatively low temporal resolution of an RGB camera. However, this method contains two limitations. First, the methods are based on a single pulse wave acquired from a face or a palm, which differs from our present study that presents a relationship between spatial differences in pulse waves and blood pressure. Second, the method response time to estimate blood pressure is slow compared to the heart rate time frames. This long response time prevents the tracking of instantaneous characteristics of pulse waves and pulse-by-pulse blood pressure measurements [9].

In this study, we propose a remote method that overcomes the two limitations mentioned above. We estimate continuous blood pressure by continually tracking spatial information of facial pulse waves. We model a relationship between spatial information of facial pulse waves and blood pressure, based on a convolutional neural network (CNN). Estimations of continuous blood pressure at a given time are based on a representation of a pulse wave on a time axis and a phase axis, and on a pseudo-continuous time variable built by an interpolation.

2. Method

The proposed method consists of two main steps. The first step is a spatial description of the facial pulse wave. The second step involves a CNN-based model of the relationship between the spatial information of facial pulse waves and blood pressure.

2.1 Step 1: Spatial Description of the Facial Pulse Wave

Fig.1 shows the data flow of the phase of a spatial description of a facial pulse wave.

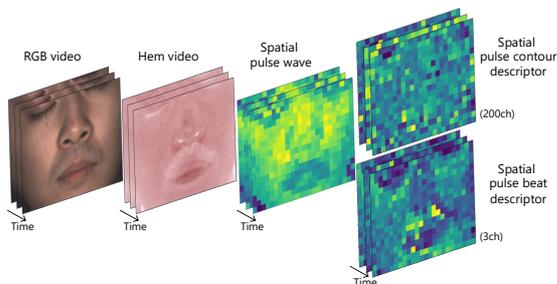


Fig 1. Dataflow of the spatial description of a facial pulse wave

This phase consists of three main processes. The first process is the construction of *spatial pulse waves* from a

facial RGB video. The second process is the construction of a *spatial pulse contour descriptor* from the *spatial pulse wave*. The third process is the construction of a *spatial pulse beat descriptor* from the *spatial pulse wave*.

A conceptual diagram of the construction of *spatial pulse waves* is shown in Fig. 2. This process is based on a remote method to extract pulse waves from RGB facial video as proposed by Fukunishi et al. [10]. The method separates a facial RGB image into intensity maps of melanin, hemoglobin (Hem), and shading (residual information). In their method, a single pulse wave is extracted from the generated hemoglobin map, by choosing a region of interest in the RGB facial video. The extracted pulse wave is denoised by the detrending method and bandpass filtering. In our method, multiple regions of interest (40 px × 40 px) are set to cover a face in a mutually exclusive and collectively exhaustive manner. Then, pulse waves corresponding to each region of interest are extracted. The extracted pulse waves are spatially reconstructed in such a way as to preserve the spatial phase relationship of the original video, constructing a *spatial pulse wave*.

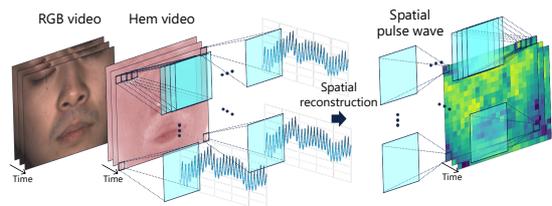


Fig 2. Conceptual diagram of the construction of a spatial pulse wave

A conceptual diagram of the construction of a *pulse contour descriptor* is shown in Fig. 3. Below, we will first describe a *pulse contour descriptor* and then describe a *spatial pulse contour descriptor*. A *pulse contour descriptor* is constructed from a pulse wave at a single coordinate of a spatial pulse wave so that it does not have spatial information. A *spatial pulse contour descriptor* is constructed from all pulse waves at all coordinates of a spatial pulse wave so that it has spatial information.

A *pulse contour descriptor* is constructed in three steps. First, a pulse wave is decomposed into unit pulse contours based on peak detection. Second, setting the start point of a pulse contour at zero phase angle, each pulse contour is separated into a time axis and a phase axis. This process collapses each pulse contour into a single phase-description at a single point of the time axis (Fig. 3). The phase resolution is set at 200 points per cycle. Third, an interpolation process is applied to the magnitudes of each phase along the time axis. This is to resample linearly pulse contours, which appear at unequal intervals. The interpolation process is performed by cubic spline interpolation at 200 Hz.

A *spatial pulse contour descriptor* is constructed in two steps. First, a *pulse contour descriptor* is extracted for each spatial coordinate of the face. Second, a *pulse*

contour descriptor for each spatial coordinate of the face is spatially reconstructed, preserving the spatial relationship of the facial pulse wave. This process results in a *spatial pulse contour descriptor* with 200 channels (phase resolution).

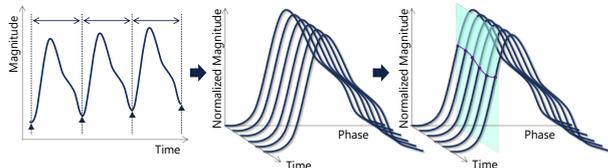


Fig 3. Conceptual diagram of a construction of a *pulse contour descriptor*. The start point of a pulse contour is set at zero phase angle.

For the construction of a *spatial pulse beat descriptor*, we will first describe a *pulse beat descriptor*, and then we will describe a *spatial pulse beat descriptor*.

The *pulse beat descriptor* is constructed via three processes. First, three pieces of information are extracted from the hemoglobin video component: a *pulse phase*, a *pulse volume*, and a *pulse interval*.

The *pulse phase* is defined as the difference between the pulse peak time of each region of interest and the average pulse peak time of all regions of interest.

The *pulse volume* is defined as the ratio between the

AC and DC components of the pulse wave.

The *pulse intervals* are defined as the time differences between the peak times of consecutive pulses.

Second, pulse phase, pulse volume, and pulse interval are resampled along the time axis. Resampling is performed by a cubic spline interpolation with a sampling rate of 200 Hz.

The *spatial pulse beat descriptor* is constructed via two processes. First, a pulse beat descriptor is extracted for each spatial coordinate of a spatial pulse wave. Second, a *pulse beat descriptor* for each spatial coordinate of the image is reconstructed to preserve the spatial relationship of spatial pulse waves. These processes result in the construction of a *spatial pulse beat descriptor*, which holds three channels corresponding to a *pulse phase*, a *pulse volume*, and a *pulse interval*.

2.2 Step 2: CNN for blood pressure training

We use a deep-learning CNN architecture based on ResNet [11] and CBAM [12]. The *spatial pulse contour descriptor* and the *spatial pulse beat descriptor*, previously defined, are the inputs. Systolic blood pressure (SBP), mean blood pressure (MBP), and diastolic blood pressure (DBP) are the outputs of the CNN (Fig. 4). Each module in the deep learning architecture is shown in Figure 5.

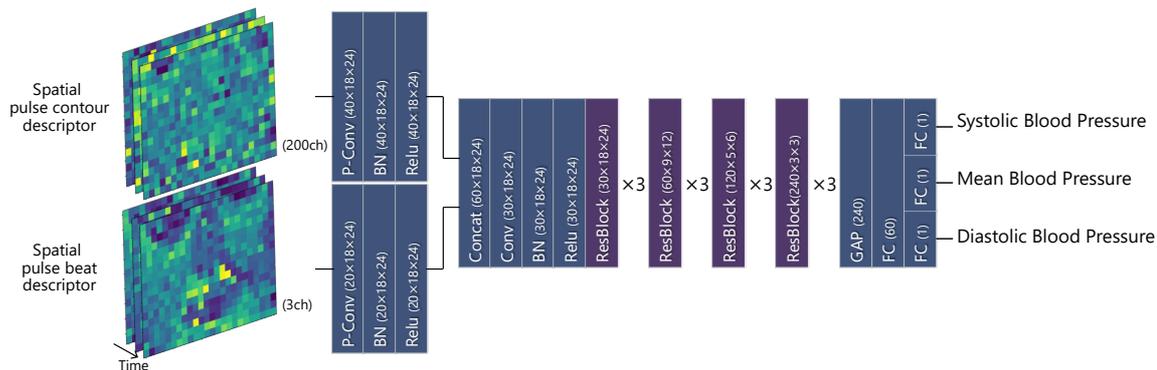


Fig 4. Deep learning architecture. Abbreviation's guide: Conv (convolutional layer), P-Conv (point-wise convolutional layer), BN (batch normalization), Concat (concatenate), GAP (global average pooling), and FC (fully connected layer). The *spatial pulse contour descriptor* and the *spatial pulse beat descriptor* are independently encoded. Then, they are concatenated and encoded. The *spatial pulse descriptor* contains a *pulse phase*, a *pulse volume*, and a *pulse interval*.

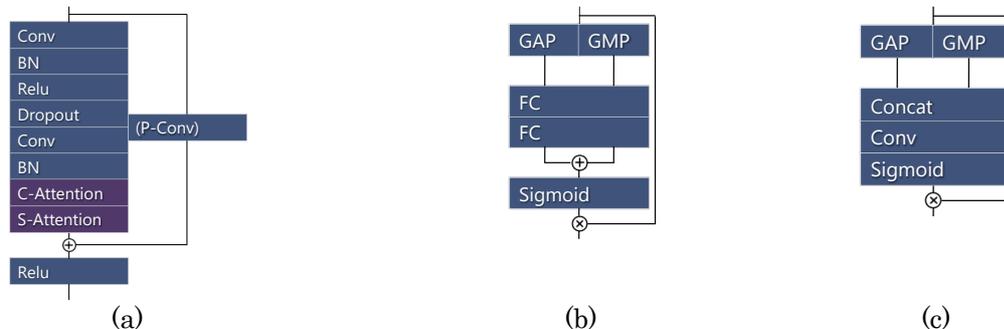


Fig 5. Architecture of each block in the architecture of deep learning. Abbreviations guide: Conv (convolutional layer), P-Conv (point-wise convolutional layer), BN (batch normalization), Concat (concatenate), GAP (global average pooling), GMP (global max pooling), and FC (fully connected layer). (a) ResBlock (b) Channel attention block (C-Attention) (c) Spatial attention block (S-Attention)

3. Experiment

In what follows, first, we describe an experiment we conducted with volunteers to construct a dataset. Second, we describe a benchmarking design. Third, we describe the training of the CNN. Finally, study the accuracy of our proposed method by comparing predictions with the ground truth

3.1 Dataset Construction

An experiment with volunteers was conducted to build a dataset to evaluate the effectiveness of the proposed method. The number of subjects was 10 (9 males and 1 female), aged 23.3 ± 1.4 years old. The experimental environment is shown in Fig. 6. The protocol used to modify the blood pressure of the volunteers during the data acquisition process was a cycle consisting of 30 seconds of resting state, up to 60 seconds of breath-holding state, and 60 seconds of resting state. This protocol was performed 3.0 ± 0.5 times for each volunteer, which produced a total of 30 measurements for all volunteers. The variation in the number of times was due to data corruption caused by an equipment malfunction. We acquired blood pressure data with a continuous monitor (Finometer MIDI, Finapres Medical Systems) attached to the left middle finger of the volunteers, and acquired face video with an RGB camera (DFK33UX174, The Imaging Source). The camera sampling rate was set at 160 Hz, and its resolution was set at $960 \text{ px} \times 740 \text{ px}$. Boxplots of the blood pressure of all measurements are shown in Fig.7.



Fig 6. Experimental environment

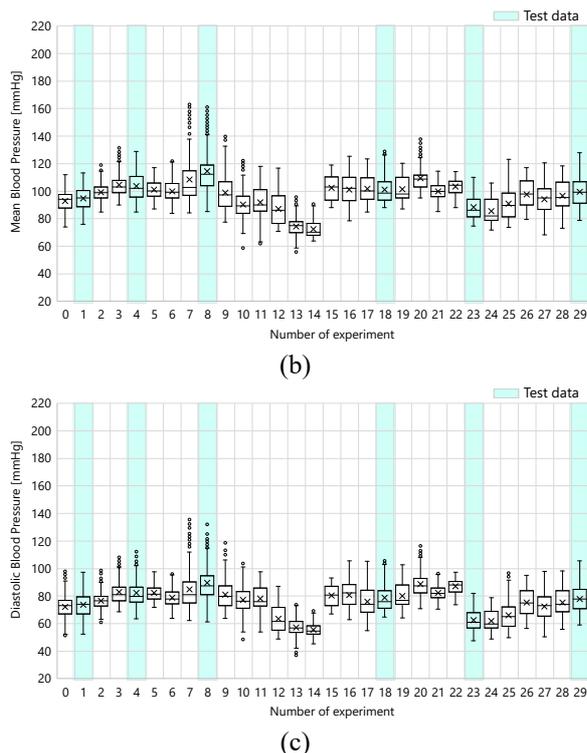
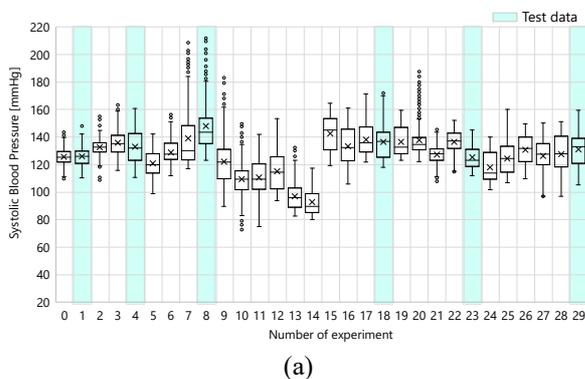


Fig. 7. Boxplots of blood pressure of constructed dataset. (a) Systolic blood pressure (b) Mean blood pressure (c) Diastolic blood pressure. The green vertical bars show test data.

3.2 Benchmarks Design

To use as a benchmark, we designed a multilayer perceptron that estimates blood pressure from a set of conventional features extracted from facial pulse waves. The multilayer perceptron is a standard model for a neural network that does not deal with spatial information. This multilayer perceptron has five intermediate layers. The number of channels of the input layer is the number of features. The number of channels at the end of the intermediate layer is half of the number of channels in the input layer. The number of channels of the intermediate layers linearly decrease from the number of channels in the input layer to the number at the end of the intermediate layer, and so on. The channels of the output layer are systolic, mean, and diastolic blood pressure. As inputs, we chose pulse wave features known to pertain to blood pressure: pulse wave contour, its second derivative, and pulse beat [2,13,14,15]. Table 1 shows the full list of features.

3.3 Training the Neural Networks

This section describes the training of the deep learning architecture for the evaluation of the proposed method. The total number of data points is 9000, [30 experiments \times each experiment's duration (150 s) \times sampling rate (200 s^{-1}). This sampling rate is the one from the *spatial pulse contour descriptor* and *spatial pulse beat descriptor*. The

dataset is divided into training and testing data in units of one experiment. We use 80% of the dataset (24 experiments) for training data and the remaining 20% of the dataset (6 experiments) for testing. Test data has different experimental data from the same individuals as training data. In test data, there is no duplication of the individuals. In training, the number of epochs is 50. The batch size is 256. The loss function is defined as the sum of the mean squared error (MSE) of systolic, mean, and diastolic blood pressure. The optimization method is the Adam Optimizer. The learning rate starts at 0.01 and it is divided by 5 every 10 epochs.

3.4 Results and Discussion

To compare the proposed method with the benchmarks, the correlation coefficient and the mean absolute error are shown in Table 2. The proposed method shows significantly superior results, with a correlation coefficient of over 50% better, and with a mean absolute error of 35% lower than the benchmark we used. In addition, the proposed method tracks ground truth blood pressure waveforms qualitatively much better than the benchmark (Fig. 8). Overall, the proposed method, under

the conditions tested, is adequate on its own merits: the correlation coefficient between predicted and ground truth MBP is 85%, with a mean absolute error of 5.4 mmHg, which is about 5% of the MBP.

Here, we discuss the limitations of the proposed method. There are two main limitations. First, the influence of individual differences on the proposed method is not sufficiently demonstrated in this study. Because in this study, the training and testing data contained information from the same individuals. Second, the data were acquired under idealized conditions. Requirements for the quality of pulse waves obtained remotely require further studies. It is well known that the quality of pulse waves obtained remotely depends on the characteristics of the camera, light source, skin, and body motion of the subjects [9,10,16,17,18]. In addition, we need to investigate the precision and robustness of each part of the proposed process, such as peak detections. Third, we heuristically hypothesized that information about blood pressure could be obtained from the pulse wave propagation on the face. A theoretical explanation of the physiological connection between the analyzed pulse waves and blood pressure is beyond the scope of the current study.

Table 1. Features for benchmark

Concept	Feature
Contour	Overall area of pulse contour
	Systolic area of pulse contour
	Diastolic area of pulse contour
	Phase of peak of pulse contour
	Diastolic area / Systolic area of pulse contour
	Width of pulse contour (10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%)
Derivative	Phase of a-peak of 2nd derivative of pulse contour
	Magnitude of a-peak of 2nd derivative of pulse contour (a)
	Phase of b-peak of 2nd derivative of pulse contour
	Magnitude of b-peak of 2nd derivative of pulse contour (b)
	Phase of c-peak of 2nd derivative of pulse contour
	Magnitude of c-peak of 2nd derivative of pulse contour (c)
	Phase of d-peak of 2nd derivative of pulse contour
	Magnitude of d-peak of 2nd derivative of pulse contour (d)
	Phase of e-peak of 2nd derivative of pulse contour
	Magnitude of e-peak of 2nd derivative of pulse contour (e)
	b / a
	c / a
	d / a
	e / a
	$(c - b) / a$
$(d - b) / a$	
$(b - c - d - e) / a$	
Pulse beat	Interbeat interval of pulse wave
	AC/DC ratio of pulse wave

Table 2. Estimation error between the proposed method and benchmarks

	Correlation coefficient			Mean absolute error [mmHg]		
	SBP	MBP	DBP	SBP	MBP	DBP
Benchmark (Pulse beat)	0.10	0.12	0.09	87	59	33
Benchmark (Derivative)	0.23	0.23	0.25	12	9.9	9.4
Benchmark (Contour)	0.37	0.36	0.38	11	10	9.6
Benchmark (All)	0.52	0.55	0.54	9.8	8.3	8.2
Proposed method	0.81	0.85	0.84	6.7	5.4	5.4

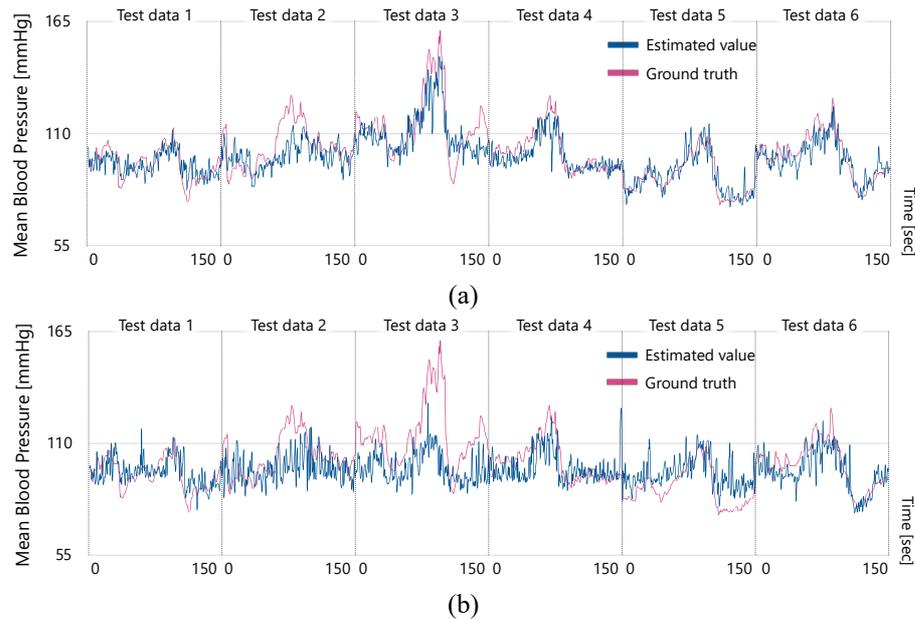


Fig 8. Accuracy: blood pressure ground truth vs. estimated value. (a) Proposed method (b) Benchmark (using all features).

4 Conclusion and future work

We proposed a remote method to estimate continuous blood pressure based on time-space information of pulse waves as observed on the face of a subject by an RGB camera. We modeled a relationship between the time-space signals and blood pressure by a convolutional neural network. A dataset was constructed to demonstrate the effectiveness of the proposed method, consisting of continuous blood pressure data and facial RGB video of ten healthy volunteers. The demonstration was performed by comparing the error and the accuracy of the proposed method with benchmarks representing conventional methods. The demonstration showed the superior performance of the proposed method compared to benchmarks, contributing to ubiquitous blood pressure monitoring

There are three major issues to be addressed in the future. The first is to increase the reliability of the demonstration by having a larger and more diverse

dataset, the second is to evaluate the robustness of the proposed method for non-restrictive environments, and the third is to generalize the proposed concept to measure other physiological quantities.

Regarding the enhancement of the reliability of the demonstration, there is a concern about the reliability of the demonstration of the proposed method in this study due to biases such as the number of subjects and the age of the subjects. To resolve this concern, it is necessary to construct a more reliable data set and demonstrate the proposed method based on it.

Regarding robustness, the method should be able to withstand a noisy, non-restrictive, or non-controlled environment. If the proposed method is not sufficiently robust in non-restrictive environments, it is necessary to consider adding features to it.

Regarding the generalization of the proposed concept for other physiological quantities, the proposed method is expected to be effective not only for the continuous blood pressure estimation task but also for other tasks such as emotion recognition. This hypothesis is based on the

plurality of pulse-wave information. The plurality of information in a pulse wave is supported by the existence of studies that estimate quantities such as stress, emotion, advertising effects, and oxygen saturation, based on pulse waves [16,17,18,19].

References

- [1] T. Nagasawa, K. Iuchi, R. Takahashi, M. Tsunomura, R. P. Souza, K. Ogawa-Ochiai, N. Tsumura, G. C. Cardoso. Blood Pressure Estimation by Photoplethysmogram Decomposition into Hyperbolic Secant Waves. *Applied Sciences*, 12(4), 1798 2022
- [2] S. G. Khalid, H. Liu, T. Zia, J. Zhang, F. Chen, D. Zheng. Cuffless Blood Pressure Estimation Using Single Channel Photoplethysmography: A Two-Step Method. *IEEE Access*, 8, 58146-58154, 2020
- [3] I. C. Jeong, J. Finkelstein. Introducing Contactless Blood Pressure Assessment Using a High Speed Camera. *Journal of medical systems*, 40(4), 1-10, 2016
- [4] X. Fan, Q. Ye, S. D. Choudhury. Robust blood pressure estimation using an RGB camera. *Journal of Ambient Intelligence and Humanized Computing*, 11, 4329-4336, 2018
- [5] P. W. Huang, C. H. Lin, M. L. Chung, T. M. Lin, B. F. Wu. Image based contactless blood pressure assessment using Pulse Transit Time. In: *IEEE International Automatic Control Conference*, pp.1-6, 2017
- [6] D. Buxi, J. M. Redoute, M. R. Yuce. A survey on signals and systems in ambulatory blood pressure monitoring using pulse transit time. *Physiological measurement*, 36(3), 1-26, 2007
- [7] N. Sugita, M. Yoshicawa, M. Abe, A. Tanaka, N. Homma, T. Yambe. Contactless Technique for Measuring Blood-Pressure Variability from One Region in Video Phethysmography. *Journal of Medical and Biological Engineering*, 39(1), 76-85, 2019
- [8] M. Rong, K. Li. A blood pressure prediction method based on imaging photoplethysmography in combination with machine learning. *Biomedical Signal Processing and Control*. 64, 2021
- [9] R. Takahashi, K. Ogawa-Ochiai, N. Tsumura. Non-contact method of blood pressure estimation using only facial video. *Artificial Life and Robotics*, 25(3), 343-350, 2020
- [10] M. Fukunishi, K. Kurita, S. Yamamoto, N. Tsumura. Non-contact video-based estimation of heart rate variability spectrogram from hemoglobin composition. *Artificial Life and Robotics*, 22(4), 457-463, 2017
- [11] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016
- [12] S. Woo, J. Park, J. Y. Lee, I. S. Kweon. Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 3-19, 2018
- [13] M. Liu, L. M. Po, H. Fu. Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative. *International Journal of Computer Theory and Engineering*, 9(3), 202-206, 2017
- [14] S. S. Mousave, M. Firouzmand, M. Charimi, M. Hemmati, M. Moghadam, Y. Ghorban. Blood pressure estimation from appropriate and inappropriate PPG signals using A whole based method. *Biomedical Signal Processing and Control*, 47, 196-206, 2019
- [15] K. Matsumura, P. Rolfe, S. Toda, T. Tamakoshi. Cuffless blood pressure estimation using only a smartphone. *Scientific reports*, 8(1), 1-9, 2018
- [16] K. Iuchi, R. Mitsuhashi, T. Goto, A. Matsubara, T. Hirayama, H. Hashizume, N. Tsumura. Stress levels estimation from facial video based on non-contact measurement of pulse wave. *Artificial Life and Robotics*, 25(3), 335-342, 2020
- [17] G. Okada, T. Yonezawa, K. Kurita, N. Tsumura. Monitoring emotion by remote measurement of physiological signals using an RGB camera. *ITE Transactions on Media Technology and Applications*, 6(1), 131-137, 2018
- [18] K. Masui, G. Okada, N. Tsumura. Measurement of advertisement effect based on multimodal emotional responses considering personality. *ITE Transactions on Media Technology and Applications*, 8(1), 49-59, 2020
- [19] A. R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M. C. Frise, P. A. Robbins, L. Tarassenko. Non-contact measurement of oxygen saturation with an RGB camera. *Biomedical optics express*, 6(9), 3320-3338, 2015