

Overall Description:

CIME PPG dataset described in "Ke Xu *et al.*, 'Photoplethysmography Motion Artifacts Removal based on Signal-Noise Interaction Modeling Utilizing Envelope Filtering and Time-Delay Neural Network,' *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3732-3744, Apr. 2020. <https://doi.org/10.1109/JSEN.2019.2960370>" (training set) and in "Ke Xu *et al.*, 'Deep Recurrent Neural Network for Extracting Pulse Rate Variability from Photoplethysmography During Strenuous Physical Exercise,' in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Nara, Japan, 2019, pp. 1-4. <https://doi.org/10.1109/BIOCAS.2019.8918711>" (testing set).

The detailed description of the dataset can be found in attached file "dataset_description.pdf".

Since the data size is beyond the maximum size required by GitHub, please go to the following Dropbox link for data files: <https://www.dropbox.com/sh/6np4q7dg9iz46ki/AACZd58eD8iVVOHIR-9vEUbRa?dl=0>.

For RESEARCH USE ONLY.

Please cite the TWO articles above if using this dataset (training and testing set) for publication.

Experiment Description:

48 adult subjects were invited for the experiments, whose ages vary from 18 to 65 (20 males and 28 females, with white, yellow and dark skin colors, with or without history of smoking). Written consent was acquired from each participant prior to the experiment sessions. This was a non-clinical study without any harming procedure and all data were collected anonymously. These experiment procedures follow strictly with the principles of the Declaration of Helsinki.

The gloves described in "Photoplethysmography Motion Artifacts Removal based on Signal-Noise Interaction Modeling Utilizing Envelope Filtering and Time-Delay Neural Network" were worn on both hands of subjects with PPG detector cuffs fixed on index fingertips. A certificated medical level pulse oximeter (KS-CM01, Li Kang Co., Ltd.) was fixed on middle finger of left hand for reference. During the whole experiments, subjects' left hands were required to keep static, while right hands act following the PC software's instructions, which include continuous movement and rest.

For subject 1-10 (training set), 10 sets of experiments were carried out for each person. For subject 11-48 (testing set), 1 set of experiments was carried out for each person. Each set of experiment contains 10 sequences. For the 1st to the 5th sequence, the movement pattern of bending index finger was required, while for the 6th to 10th sequence, the movement pattern of horizontal finger-axial moving was required. The movements are instructed to be carried out in large and small amplitudes, high and low frequencies, whereas the exact amplitudes and frequencies are not given and up to the individuals themselves. Amplitudes and frequencies variation were taken into consideration because we want to include different levels of noise intensity. The two movement patterns we selected are those proven to have the most significant influences on PPG Signal-Noise Ratio (SNR).

The detailed instructions for each sequence are shown in Fig. 1. The length for each set of 10-sequences-included experiment is about 32 minutes. There is also a compulsory interval of at least one hour between different sets, during which all the experiment setups including the PPG cuffs are rearranged, in order to take the sensor contact pressure and personal physical situation fluctuations such as heart rate variation and blood pressure alternation into consideration.

For subject 1-10 each, the data length of 10 sets of experiment together is about 320 minutes, and for

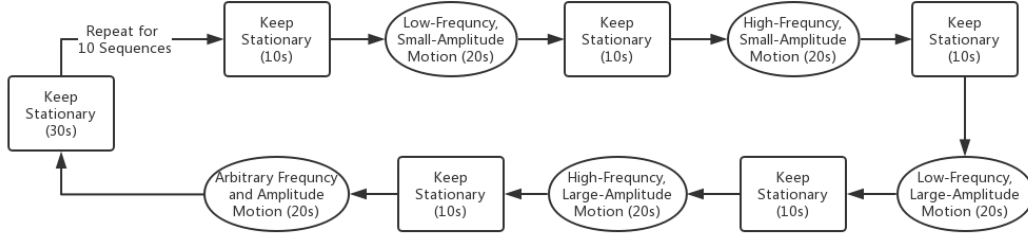


Fig. 1. Detailed instructions for each experiment sequence. The proposed sequence includes five 20-seconds-motion-instructions with different frequencies and amplitudes in order to take different PPG SNR into consideration. Intervals for resting are arranged between motions. For one set of experiment, 10 sequences above are required, the 1st to the 5th of which demands motion of bending the index fingers, and the 6th to the 10th demands motion of horizontal finger-axial moving. 10 sets of experiments in total are carried out on each subject.

the training dataset combining 10 subjects' experiments, the total data length is about 53 hours, at the sample rate of 200 Sample/s. For subject 11-48 each, the data length of 1 set of experiment is about 32 minutes, and for the testing dataset combining 38 subjects' experiments, the total data length is about 20.3 hours, at the sample rate of 200 Sample/s.

Data Format:

The dataset contains two parts: training set and testing set. **Both have been down-sampled to 100 Samples/s.**

train_all_noise.mat:

- (1) noise_input_all: 5*1000 matlab cell matrix, each cell contains 30s data (100Sps) with 3 rows, corrupted ppg in 1st row, x-axis accelerometer in 2nd row, y-axis gyroscope in 3rd row
- (2) noise_output_all: 5*1000 matlab cell, each cell contains 30s data (100Sps) with 1 row, clean ppg in 1st row

Each column of the matrix represents for 1 sequence, and 1 set of experiment contains 10 sequences (10 columns). For training, 100 sets (from 10 subjects) are used.

test_all_noise.mat:

- (1) noise_input_all: 5*380 matlab cell matrix, each cell contains 30s data (100Sps) with 3 rows, corrupted ppg in 1st row, x-axis accelerometer in 2nd row, y-axis gyroscope in 3rd row
- (2) noise_output_all: 5*380 matlab cell, each cell contains 30s data (100Sps) with 1 row, clean ppg in 1st row

Each column of the matrix represents for 1 sequence, and 1 set of experiment contains 10 sequences (10 columns). For testing, 38 sets (from 38 subjects) are used.

NOTE: The length of each column combined is 30*5=150s, thus for 1 set of experiment which contains 10 columns the total length is 1500s (25min). It is shorter than the original experiment length (31min) because we removed the time periods at the beginning and end of each sequence where the subjects were kept still and no movement was carried out. Therefore, we suggest not to concatenate different sequences (columns) because they are not continuous. However, the combination within a single sequence (column) is OK.

Please contact us with email: kexu18@fudan.edu.cn for any question.