Supplementary Annotation Dataset

Optimal Preprocessing of Raw Signals from Reflective Mode Photoplethysmography in Wearable Devices

Florian Wolling, Sudam Maduranga Wasala, and Kristof Van Laerhoven
Universit of Siegen, Germany

January 03, 2022

1 General Information

1.1 Contact

- Florian Wolling — florian.wolling@uni-siegen.de
- Sudam Maduranga Wasala
- Kristof Van Laerhoven — kvl@eti.uni-siegen.de

1.2 Web Pages

- Supplementary Annotation Dataset [**1**] — https://ubicomp.eti.uni-siegen.de/home/datasets/embc21/
- Original Dataset of Raw PPG [**2**] — https://doi.org/10.1016/j.dib.2019.105044
- Florian Wolling — https://ubicomp.eti.uni-siegen.de/home/team/fwolling.html.en
- Kristof Van Laerhoven — https://ubicomp.eti.uni-siegen.de/home/team/kristof.html.en

1.3 Disclaimer

The provided, publicly available annotation data supplement the original recordings of Biagetti et al. [**2**], but do not themselves contain them. You may use the annotations, manually set by an expert rater, as a reference for scientific, non-commercial purposes, provided that you give credit to the owners when publishing any work based on it. We would also be very interested to hear back from you if you use our annotations in any way and are happy to answer any questions or address any remarks related to it.

1.4 Citation

When using our annotations for your work, please consider providing the following two references in your bibliography, since our work [**1**] is based on the original dataset of Biagetti et al. [**2**].

[**1**] "Optimal Preprocessing of Raw Signals from Reflective Mode Photoplethysmography in Wearable Devices", Florian Wolling, Sudam Maduranga Wasala, and Kristof Van Laerhoven. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC 2021, Virtual Event, November 2021, IEEE, 2021.
https://doi.org/10.1109/EMBC46164.2021.9630955

[**2**] "Dataset from PPG wireless sensor for activity monitoring", Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, Leonardo Saraceni, Andrea Tiranti, Claudio Turchetti. Data in Brief, Volume 29, 105044, 2020, ISSN 2352-3409.
https://doi.org/10.1016/j.dib.2019.105044

2 Purpose

In our research [1], we highlight the importance of benchmarking algorithms on actual raw signals from reflective mode photoplethysmography (PPG). To demonstrate and evaluate the impact of preprocessing on pulse peak positions and the performance of peak detection algorithms, an expert rater manually annotated the peak labels to serve as ground truth for the publicly available dataset of Biagetti et al. [2]. We encourage researchers to use the dataset of raw PPG signals in combination with the supplementary annotations from our research [1] to benchmark their own algorithms as well as machine learning approaches. The annotation files of the 21806 diastolic pulse onset labels, provided as serialized Python Pickle (*.p) and Comma-Separated Values (*.csv), are available for download.

More information about the annotations and our research are available on the web page:
https://ubicomp.eti.uni-siegen.de/home/datasets/embc21/

3 Dataset

3.1 File Structure

The *.zip archive *ppgopt_annot.zip* encloses the two folders **csv** and **pkl** which contain the identical annotation data, the manually determined pulse peak positions for the raw PPG recordings of the dataset [2]. The data have, however, been exported to two different file types, the preferable, more comfortable Python Pickle (*.p) and the general, common standard, but cumbersome Comma-Separated Values (*.csv). The archive also contains the *README.pdf* documentation in your hands. Both folders again contain seven sub-folders **S1** to **S7** associated with the seven subjects. The archive's file structure is illustrated below.

- *ppgopt_annot.zip*
  - csv
    - S1
    - ...
    - S7
      - s7_rest1_atf.csv
      - s7_rest1_pks.csv
      - ...
      - s7_squat1_atf.csv
      - s7_squat1_pks.csv
      - ...
      - s7_step1_atf.csv
      - s7_step1_pks.csv
      - ...
  - pkl
    - S1
    - ...
    - S7
      - s7_rest1.p
      - ...
      - s7_squat1.p
      - ...
      - s7_step1.p
      - ...
  - README.pdf

3.1.1 Python Pickle (*.p)

The serialized Python Pickle files (*.p) allow comfortable access to the annotations. Located in the folder pkl, the seven subjects' individual sub-folders **S1** to **S7** each contain files labeled according to the schema <subject>_<exercise><run>.p. The annotations can easily be read in using Python. A simple code example, to obtain the pulse peak labels and artifact intervals, is given below:

```python
import pickle
# Load the annotation data of subject S1, squat 2.
file = "./pkl/S1/s1_squat2.p"
data = pickle.load(open(str(file), "rb"))
# Obtain the annotations and artifact intervals
peaks = data["peaks"]  # The peak labels.
artifacts = data["artifacts"]  # The artifact intervals.
```

3.1.2 Comma-Separated Values (*.csv)

The Comma-Separated Values files (`*.csv`) allow access to the reference data with any preferred software. Located in the folder `csv`, the seven subjects' individual sub-folders **S1** to **S7** each contain files labeled according to the schema `<subject>_<exercise><run>_<type>.p`. The `type` here defines whether the file contains a list of the pulse peak labels `pks` or the artifact intervals `atf`. The structure of the contained information is kept simple and there is no header. In case of `pks`, the rows contain only single integer values, the indices of identified pulse peaks in the original time series. In case of `atf`, the file is either empty, if no artifact intervals are specified, or contains rows with two columns, separated through a single semicolon ';', defining begin and end of the artifact intervals.

3.2 Details

Below, Table 1 provides the dataset summary, specifying length (T), percentage of excluded intervals (MA), number of identified peaks (PKS), and average heart rate (HR) for each individual recording of the subjects (S) with the exercises (EX) and the five runs (SUBSET).

Table 1: Dataset summary. In total 104 of 105 time series, 278 of 286 min (97.3 %) are annotated with 21 806 peak labels, only 7.73 min do not contain distinguishable signals or are considerably affected by motion artifacts and hence excluded. 88 of 105 time series are entirely labeled (green). Exercise squat 3 of subject 5 (red) only contains noise and is hence excluded. S: subject; EX: exercise; T: length; MA: excluded part; PKS: number of labeled peaks; HR: average heart rate.

S	EX	SUBSET																			
		1				2				3				4				5			
		T (s)	MA (%)	PKS	HR (bpm)	T (s)	MA (%)	PKS	HR (bpm)	T (s)	MA (%)	PKS	HR (bpm)	T (s)	MA (%)	PKS	HR (bpm)	T (s)	MA (%)	PKS	HR (bpm)
S1	rest	593	5.87	649	69.13	532	3.15	592	68.12	577	5.15	603	65.39	570	4.58	738	80.09	1000	6.83	896	57.35
	squat	62	–	104	101.4	59	–	101	103.93	69	–	112	99.0	61	1.35	103	103.52	60	–	102	103.41
	step	84	–	160	115.65	77	–	144	114.44	68	4.69	119	109.68	106	–	198	113.01	108	18.71	171	115.27
S2	rest	648	9.53	651	66.0	554	7.79	550	63.49	556	17.48	485	58.98	686	4.38	856	77.41	519	4.56	610	72.38
	squat	44	–	59	81.17	43	–	72	100.88	40	–	63	95.42	47	–	76	97.7	43	–	69	96.79
	step	63	–	94	92.19	75	–	115	93.69	72	–	108	90.52	95	–	145	92.14	94	–	147	95.05
S3	rest	283	–	319	67.82	274	–	316	69.57	226	–	272	72.68	262	–	308	70.56	278	–	339	73.63
	squat	49	–	78	95.92	43	–	70	98.89	45	–	71	95.3	46	–	75	99.07	49	–	83	101.26
	step	61	–	92	91.94	59	–	91	93.41	49	–	79	97.93	52	–	80	92.81	51	–	78	92.99
S4	rest	265	–	290	65.89	278	–	307	66.49	271	–	297	65.87	287	–	319	66.88	261	–	294	67.9
	squat	52	–	68	79.25	40	–	56	83.04	39	–	53	81.52	38	–	50	79.82	44	–	59	81.36
	step	45	–	59	79.87	49	–	64	78.37	55	–	76	82.59	54	–	74	82.4	66	–	90	82.47
S5	rest	287	–	404	85.85	289	–	413	86.63	291	–	372	77.27	288	–	381	79.94	287	–	363	76.7
	squat	32	–	53	99.15	37	–	56	90.92	61 s	100.00 %	0	/	55	–	79	87.4	61	–	89	89.38
	step	51	–	97	114.35	46	–	83	110.95	47	–	87	113.44	51	–	93	111.62	48	–	94	118.48
S6	rest	278	–	354	76.76	286	–	375	79.06	278	–	341	74.02	278	–	356	77.14	282	–	358	76.71
	squat	42	–	62	89.39	47	–	69	88.79	51	–	78	91.95	49	–	76	93.02	48	–	77	95.1
	step	61	–	93	91.37	69	–	99	87.11	66	–	99	90.76	66	–	104	93.93	63	–	95	91.47
S7	rest	329	–	483	88.4	311	–	425	82.64	287	0.76	384	80.83	293	–	370	76.18	290	0.76	361	75.44
	squat	56	6.52	103	120.89	49	–	96	119.38	48	–	98	126.14	58	–	115	122.57	56	–	85	91.67
	step	55	–	117	130.2	59	–	135	139.52	43	–	99	141.22	50	–	116	140.14	48	–	120	150.11