

On indirect assessment of heart rate in video

Mikhail Kopeliovich

Konstantin Kalinin

Yuriy Mironenko

Mikhail Petruschan

Southern Federal University, Center of Neurotechnologies
Rostov-on-Don, Russian Federation

kop@km.ru

Abstract

Problem of indirect assessment of heart rate in video is addressed. Several methods of indirect evaluations (adaptive baselines) were examined on Remote Physiological Signal Sensing challenge. Particularly, regression models of dependency of heart rate on estimated age and motion intensity were obtained on challenge's train set. Accounting both motion and age in regression model led to top-quarter position in the leaderboard. Practical value of such adaptive baseline approaches is discussed. Although such approaches are considered as non-applicable in medicine, they are valuable as baseline for the photoplethysmography problem.

1. Introduction and prior work

RePSS (The 1st challenge on Remote Physiological Signal Sensing) was held in conjunction with CVPR 2020 [1–4]. The challenge aimed at comparing accuracy of different methods of heart rate assessment in video with stationary and moving humans. Common approaches on heart rate estimation in video are mostly based on remote photoplethysmography (rPPG) which is analysis of skin color changes caused by heart beat [5, 6] or ballistocardiography which is detection of tiny motions caused by heart pulsation [7].

We, however, tried to indirectly “guess” a heart rate. Our guess was based on assessment of age and motion in the video, without using of any PPG-related signal at all. We called this approach “adaptive baseline”.

The inverse problem was, to great extent, solved. Heart rate (HR) as a proxy variable for evaluation of energy expenditure is a well-known and widely used technique [8, 9]. This technique typically involves the use of other factors, such as body weight index (BWI), gender or age [10, 11], to produce more accurate estimation. As long as such estimation is accurate enough for the practical usage [12], the following question appears: is it possible to invert this scheme to evaluate HR itself? Is it possible to combine the information about BWI, age, gender, physical activity or other

factors, and produce good HR guess?

It turns out that our adaptive baseline was actually better than 3/4 of other participants' best efforts. It is a matter of discussion whether the baseline HR estimations are of any practical value, even if formally accurate enough.

2. The challenge description

The RePSS dataset is divided into training and testing sets. The training set consists of randomly cut 10-second clips from 500 randomly selected videos of VIPL-HR V2 dataset [1]. Up to date, this dataset is poorly described by the authors. Regarding the training set, it contains 500 folders of 10 sec videos and metadata. In most cases, there are five videos per folder, cut from corresponding video files of the VIPL-HR V2 dataset, resulting in 2498 videos. Metadata of a single ground truth HR value and average value of frame rate correspond to each video. Authors state that they used a Realsense F200 camera to record videos with 960×720 resolution [1]. However, the resolutions of the training set videos vary. The scene backgrounds are non-monotonous and sometimes dynamic with presence of other people behind the subject. Subjects in the dataset are of different age (including children and elderly) and gender. Total number of subjects is 500.

The testing set contains 1000 video files of 10 sec videos from 200 subjects (100 from VIPL-HR V2 and 100 from OBF Database [2]). The extra complexity is that the areas of the eyes and mouth areas are covered with mosaics. Instead, the authors provided the facial bounding box location and 68 facial landmarks.

The RePSS authors selected the mean average error (MAE) as primary metric to compare methods of HR evaluation. Two additional metrics were used: root mean square error (RMSE) and Pearson correlation coefficient (R).

3. Methods

In rPPG, average HR value can be used as baseline. It is essentially a prior knowledge about HR distribution on a

training set for a proposing rPPG method. Average HR is the simplest form of such prior knowledge. More sophisticated forms may involve additional information such as estimated age and gender of a person in video, type of activity, and others [13]. These additional factors can adapt baseline to particular conditions observable in video under analysis. In particular, we studied relations of HR on estimated age of a person and on person's motion intensity in challenge's training dataset. Further, we estimated age and motion activity in videos from test set and applied prior knowledge on HR dependencies to obtain baseline predictions.

We consider four methods for heart rate estimation: Baseline by Constant HR (BC), Baseline by Motion (BMotion), Baseline by Age (BAge), and Baseline by Age and Motion (BAM).

BC represents a constant function returning the same HR value C for any input. According to [13], normal heart rate for healthy adult at rest could be estimated as $C = 70$ bpm. However, regarding RePSS, the average HR on the training set was 87.95 bpm. We rounded this value down ($C = 87$ bpm) and used it as BC output. The rounding down was made both for sake of simplicity and basing on testing set analysis: it is half composed of videos from OBF Database, where, for some sessions, subjects are expected to sit still before the physical exercises [2]. Due to relatively high HR on the training set, we assume that average HR on testing set would be slightly lower because of before-exercise part of the OBF.

BMotion is based on HR prediction by motion amplitude. Unlike ballistocardiography [7], BMotion method estimates HR by an average motion amplitude (AMA), which is a single value generated per video:

$$AMA = \frac{1}{F - \delta} \sum_{f=\delta+1}^F \left| \frac{y_c^f}{H_{face}^f} - \frac{y_c^{f-\delta}}{H_{face}^{f-\delta}} \right|, \quad (1)$$

where f is a frame index, F is a total frame count, y_c^f is a vertical coordinate of facial rectangle in pixels, H_{face}^f is a height of facial rectangle in pixels, δ ($=10$) is an offset to calculate motion amplitude. The motion amplitude is estimated in each frame as displacement of relative vertical coordinate of facial rectangle after δ frames. We didn't exploit horizontal coordinate of facial rectangle due to noisy value of width of the detected rectangle caused by head rotations. To detect facial rectangle, we used Multitask Cascaded Convolutional Networks [14] implemented on Python by Centeno [15].

Once calculated, the AMA values and corresponding actual HR values are used as inputs for linear regression algorithm. The resulting coefficients of approximation line are used to generate HR predictions.

BAge is based on age estimation. Age was estimated on cropped facial image as one of 26 classes corresponding to a range of 15 to 40 years by Consistent Rank Logits framework [16] implemented on Python by [17]. Age values were calculated for 10 evenly selected frames in video; the overall age was estimated by averaging their 5 median values.

As for BMotion, we used linear regression to generate HR predictions.

BAM combines the previous two approaches. First, BAge is used to train the basic model of linear regression providing array of \hat{y}^{BAge} estimations on the training set. Second, the relative errors $(y_i - \hat{y}_i^{BAge}) / \hat{y}_i^{BAge}$ are used as ground truth data to train another linear regression model with the corresponding AMA values on input. Both the models applied similarly on the testing set, then the outputs of the second model are back-scaled to heart rate values: $\hat{y}_i^{BAge} + \hat{y}_i^{BAge} \hat{y}_i^{BAM}$, where \hat{y}^{BAM} are predictions of the second model.

4. Results and Discussion

RePSS challenge was one of first attempts to evaluate and compare different approaches of remote HR estimation on moving person in standardized conditions. The Table 1 provides the challenge leaderboard along with our proposed baselines.

BC is the simplest method for indirect heart rate assessment. Surprisingly, it provided lower MAE and RMSE values than one third of the all other methods. Due to zero variance of BC's answers its R value is undefined (set to zero). In order to calculate R value, one can add a slight random number to each answer of BC; this procedure wouldn't significantly change MAE or RMSE.

The adaptive baselines, BMotion, BAge and BAM, further improved the result leading to the 6th place. The error of BAM is close to 4th and 5th places (less than 1% difference).

Top three methods demonstrated MAE of $\approx 7-9$ bpm, while the error of most methods was above 12 bpm. They are not far beyond even the simple average BC method with MAE of ≈ 13.25 bpm. Adaptive Baseline approach decreases MAE to ≈ 12.48 bpm.

4.1. Possible Enhancements of the Baseline Method

Even most sophisticated of our baselines, BAM, only uses age and motion assessments. Moreover, it uses very simple form of motion assessment and combines estimations in the very simple, linear way. Considering additional factors and more sophisticated combination of them would lead to even better estimations.

Obvious additional factors to consider is gender and body weight index (BWI), as long as they often used in HR-

| # or method | MAE, bpm | RMSE, bpm | R |
|----------------|--------------|--------------|-------------|
| 1 | 6.94 | 10.68 | 0.75 |
| 2 | 7.92 | 14.38 | 0.59 |
| 3 | 8.95 | 14.16 | 0.54 |
| 4 | 12.39 | 16.09 | 0.23 |
| 5 | 12.46 | 16.20 | 0.19 |
| BAM | 12.49 | 15.84 | 0.19 |
| BAge | 12.51 | 15.85 | 0.18 |
| 7 | 12.54 | 16.09 | 0.10 |
| 8 | 12.55 | 16.06 | 0.10 |
| 9 | 12.58 | 15.97 | 0.10 |
| BMotion | 12.68 | 15.93 | 0.10 |
| 10 | 12.74 | 16.13 | 0.10 |
| 11 | 12.74 | 16.35 | 0.22 |
| 12 | 12.91 | 16.54 | 0.08 |
| 13 | 12.91 | 16.57 | 0.15 |
| 14 | 12.91 | 16.57 | 0.15 |
| 15 | 13.02 | 17.30 | 0.07 |
| BC | 13.26 | 16.29 | 0.00 |
| 16 | 13.26 | 16.29 | 0.00 |
| 17 | 13.29 | 16.75 | 0.02 |
| 18 | 13.40 | 17.87 | 0.28 |
| 19 | 13.55 | 17.46 | 0.16 |
| 20 | 13.63 | 16.65 | 0.06 |
| 21 | 14.37 | 18.75 | 0.11 |
| 22 | 14.51 | 18.97 | 0.11 |
| 23 | 14.76 | 19.11 | 0.04 |
| 24 | 15.69 | 19.74 | 0.01 |
| 25 | 20.12 | 25.55 | -0.02 |

Table 1: Leaderboard of the RePSS competition. Results are sorted by MAE, which was a primary metric of the challenge. Proposed baselines are in bold.

based techniques of indirect energy expenditure measurements. Some better form of the motion assessment should be considered, preferably one which correlates with energy expenditure. While BWI can hardly be determined from the facial video of the person, some signs of obesity or underweight still may be extracted.

Other source of inspiration is HR-based methods of indirect measurement of oxygen consumption [12, 18] and blood pressure [19, 20]. According to the inverse problem approach, they could be useful for the HR estimation. We assume that oxygen consumption and blood pressure themselves could be assessed via analysis of the average skin tone.

Third group includes features like pose, facial expression, speaking and others – factors with uncertainly related to the heart rate, which, however, may be easily extracted from the video using already existing techniques. There-

fore, they may be easily, almost for granted, tested for the applicability.

Finally, obvious enhancements is to use some simple neural network to combine all the pre-extracted factors into HR prediction.

4.2. Dataset Bias

Some factors could be dataset-specific. Particularly, if average HR is significantly different for different datasets (for example, children dataset concatenated with the adults dataset), some integral characteristics of these subsets (like different backgrounds) could be recognized and, thus, HR prediction baseline could be adapted to the subset’s average. Although such accounting of all mentioned integral characteristics will likely lead to better prediction of HR and higher positions in challenges leaderboards, it doesn’t look like a way to create practically-valuable rPPG method. Moreover, it even may be done unintentionally.

4.3. Concerns

If adaptive baseline methods could be capable to produce HR predictions close to real values basing on subject’s appearance parameters is a matter of further studies. However, if they do, they will likely fail to predict HR deviations from values typical for recognized subject’s appearance. This is due to these methods do not extract pulse signal or its derivatives from data and mostly exploit prior knowledge on HR distribution over subject’s appearance. Such unrecognizable deviations are of prime importance in medical applications, as they often indicate abnormal health state and necessity of medical assistance.

End-to-end deep models become popular in rPPG. In order to determine practical value of such models they shouldn’t be considered as black boxes generating good or bad predictions of HR values. Analysis should be performed to determine what kind of features are inferring in end-to-end nets: are they appearance-related features or pulse signal-related features. If they are appearance features than method is doing adaptive baseline calculation which has minor value in medical tasks.

4.4. Indirect assessment of Heart Rate Variability

While we failed to find any attempts to solve the problem of indirect estimation for the HR itself, there are lot of researches of this kind for the heart rate variability (HRV). Influence of age, gender, BWI (especially in cases of obesity and underweight) and other factors on HRV is a field of ongoing studies [21–27]. While no ready-made indirect assessment method was proposed yet, strong relations were found. Thus, the same indirect approach for measuring HRV can also be used. It will be susceptible to the same flaws.

Acknowledgments. The project is supported by the Russian Ministry of Science and Higher Education in the framework of Decree No. 218, project No. 2019-218-11-8185 “Creating a software complex for human capital management based on neurotechnologies for enterprises of the high-tech sector of the Russian Federation” (Internal number HD/19-22-NY).

5. Conclusion

The adaptive baseline for the HR assessment, based solely on the analysis of the subject appearance, is proposed and discussed. Even linear regression model of the dependence of heart rate on age and intensity of movements fell into the top 25% of the RePSS competition leaderboard. We consider such indirect assessment as valuable baseline for photoplethysmography problem, while non-applicable for medical applications. We outlined our concern about possible implicit indirect HR assessment in current end-to-end deep networks, which merits further study.

References

- [1] X. Li, H. Han, H. Lu, X. Niu, Z. Yu, A. Dantcheva, G. Zhao, and S. Shan, “The 1st Challenge on Remote Physiological Signal Sensing (RePSS),” *arXiv*, 2020.
- [2] X. Li, I. Alikhani, J. Shi, T. Seppanen, J. Junntila, K. Majamaa-Voltti, M. Tulppo, and G. Zhao, “The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection,” in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 242–249, May 2018.
- [3] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, “Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 151–160, Oct 2019.
- [4] X. Niu, S. Shan, H. Han, and X. Chen, “Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation,” *IEEE Transactions on Image Processing*, 10 2019.
- [5] W. Verkruyse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” *OPT EXPRESS*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [6] D. McDuff and E. Blackford, “iphys: An open non-contact imaging-based physiological measurement toolbox,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6521–6524, IEEE, 2019.
- [7] G. Balakrishnan, F. Durand, and J. Guttag, “Detecting pulse from head motions in video,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3430–3437, 2013.
- [8] H. J. G. Emons, D. C. Groenenboom, K. R. Westerterp, and W. H. M. Saris, “Comparison of heart rate monitoring combined with indirect calorimetry and the doubly labelled water ($^2\text{H}_2$ ^{18}O) method for the measurement of energy expenditure in children,” *European Journal of Applied Physiology and Occupational Physiology*, vol. 65, pp. 99–103, 1992.
- [9] Z. Yu, E. Volgyi, R. Wang, A. Ember, P. Wiklund, M. Alen, F. Tylavsky, and S. Cheng, “Comparison of heart rate monitoring with indirect calorimetry for energy expenditure evaluation,” *Journal of Sports and Health Science*, vol. 1, pp. 178 – 183, 07 2012. Children’s Physical Activity and Health: Physical activity, physical fitness, diet and the health in young people.
- [10] L. Dugas, J. Goedecke, T. Noakes, H. Hiiloskorpi, R. Laukkanen, L. van der Merwe, and E. Lambert, “Prediction of energy expenditure from heart rate monitoring during submaximal exercise,” *Journal of sports sciences*, vol. 23, pp. 289–97, 04 2005.
- [11] K. L. Rennie, S. J. Hennings, J. Mitchell, and N. J. Wareham, “Estimating energy expenditure by heart-rate monitoring without individual calibration,” *Medicine Science in Sports Exercise*, vol. 33, 2001.
- [12] F. T. Ltd., “An energy expenditure estimation method based on heart rate measurement,” 2012.
- [13] G. Quer, P. Gouda, M. Galarnyk, E. Topol, and S. Steinhubl, “Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, bmi, and time of year: Retrospective, longitudinal cohort study of 92,457 adults,” *PLOS ONE*, vol. 15, p. e0227709, 02 2020.
- [14] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Processing Letters*, vol. 23, pp. 1499–1503, Oct 2016.
- [15] I. d. P. Centeno, “MTCNN face detection implementation for TensorFlow, as a PIP package.” <https://github.com/ipazc/mtcnn>, 2019.
- [16] W. Cao, V. Mirjalili, and S. Raschka, “Consistent rank logits for ordinal regression with convolutional neural networks,” *ArXiv*, vol. abs/1901.07884, 2019.
- [17] mshehrozsajjad (nickname on GitHub), “Age and gender classification is a dual-task of identifying the age and gender of a person from an image or video.” <https://github.com/mshehrozsajjad/Age-Classification>, 2019.
- [18] C. S. Peserico, P. V. Mezzaroba, G. A. Nogueira, S. M. F. d. Moraes, and F. A. Machado, “Comparison between direct and indirect methods for the determination of the maximal oxygen uptake in female runners (comparação entre os métodos direto e indireto de determinação do consumo máximo de oxigênio em mulheres corredoras),” *Revista Brasileira de Medicina do Esporte*, vol. 17, pp. 270 – 273, 08 2011.

- [19] K. Matsumura, P. Rolfe, S. Toda, and T. Yamakoshi, "Cuffless blood pressure estimation using only a smartphone," *Scientific Reports*, vol. 8, May 2018.
- [20] G. Rogers and T. Oosthuysen, "A comparison of the indirect estimate of mean arterial pressure calculated by the conventional equation and calculated to compensate for a change in heart rate," *International Journal of Sports Medicine*, vol. 21, pp. 90–95, 2000.
- [21] J. B. Carter, E. W. Banister, and A. P. Blaber, "The effect of age and gender on heart rate variability after endurance training," *Med. Sci. Sports Exerc.*, vol. 35, pp. 1333–1340, 2000.
- [22] A. Plaza-Flórida, J. H. Migueles, J. Mora-González, P. Molina-García, M. Rodríguez-Ayllón, C. Cadenas-Sánchez, I. Esteban-Cornejo, S. Navarrete, R. María Lozano, N. Michels, J. Sacha, and F. B. Ortega, "The role of heart rate on the associations between body composition and heart rate variability in children with overweight/obesity: The activebrains project," *Frontiers in Physiology*, vol. 10, p. 895, 2019.
- [23] J. B. Carter, E. W. Banister, and A. P. Blaber, "The effect of age and gender on heart rate variability after endurance training," *Medicine and science in sports and exercise*, vol. 35, pp. 1333–40, 2003.
- [24] R. Schroeder, A. Peters, and A. Heitmann, "Short-term heart rate variability—influence of gender and age in healthy subjects.," *Medicine and science in sports and exercise*, vol. 10, 2015.
- [25] K. Karason, H. Mølgaard, J. Wikstrand, and L. Sjöström, "Heart rate variability in obesity and the effect of weight loss," *The American Journal of Cardiology*, vol. 83, no. 8, pp. 1242 – 1247, 1999.
- [26] A. I. Triggiani, A. Valenzano, V. Trimigno, A. Di Palma, F. Moscatelli, G. Cibelli, and G. Messina, "Heart rate variability reduction is related to a high amount of visceral adiposity in healthy young women," *PLOS ONE*, vol. 14, pp. 1–12, 09 2019.
- [27] A. Plaza-Flórida, J. H. Migueles, J. Mora-González, P. Molina-García, M. Rodríguez-Ayllón, C. Cadenas-Sánchez, I. Esteban-Cornejo, S. Navarrete, R. María Lozano, N. Michels, J. Sacha, and F. B. Ortega, "The role of heart rate on the associations between body composition and heart rate variability in children with overweight/obesity: The activebrains project," *Frontiers in Physiology*, vol. 10, p. 895, 2019.