

Covariance intersection to improve the robustness of the photoplethysmogram derived respiratory rate

Jia Zhang, Gaetano Scebba and Walter Karlen, *Senior Member, IEEE*

Abstract—Respiratory rate (RR) can be estimated from the photoplethysmogram (PPG) recorded by optical sensors in wearable devices. The fusion of estimates from different PPG features has led to an increase in accuracy, but also reduced the numbers of available final estimates due to discarding of unreliable data. We propose a novel, tunable fusion algorithm using covariance intersection to estimate the RR from PPG (CIF). The algorithm is adaptive to the number of available feature estimates and takes each estimates’ trustworthiness into account. In a benchmarking experiment using the CapnoBase dataset with reference RR from capnography, we compared the CIF against the state-of-the-art Smart Fusion (SF) algorithm. The median root mean square error was 1.4 breaths/min for the CIF and 1.8 breaths/min for the SF. The CIF significantly increased the retention rate distribution of all recordings from 0.46 to 0.90 ($p < 0.001$). The agreement with the reference RR was high with a Pearson’s correlation coefficient of 0.94, a bias of 0.3 breaths/min, and limits of agreement of -4.6 and 5.2 breaths/min. In addition, the algorithm was computationally efficient. Therefore, CIF could contribute to a more robust RR estimation from wearable PPG recordings.

I. INTRODUCTION

Respiratory rate (RR) is an essential vital sign to assess the medical condition of patients and abnormal RR is an important predictor of serious illness [1]. Continuous monitoring of RR and RR trend changes can detect abnormal events among general ward patients and enable early interventions [2]. Commonly used sensing methods in the clinic, such as capnometry and spirometry, are based on the analysis of ex- and inhaled air (i.e. gas flow and composition changes), but are cumbersome to wear, obstructive, or subject to strong artefacts if the environment is not controlled. Wearable devices show great potential for unobstructive and continuous monitoring and could overcome some of these limitations.

Objectively assessed RR with mobile sensors shows potential for improving the diagnosis of acute lower respiratory infections at the point-of-care [3]. RR can be estimated by analyzing the photoplethysmogram (PPG), which is increasingly available on wearable devices. The PPG waveform is known to have multiple modulations induced by respiration, such as respiratory-induced intensity (RIIV), amplitude (RIAV), frequency (RIFV) [4], width (RIWV) [5], and slope transit time variation (RISV) [6]. The current most cited benchmark method for RR estimation from the PPG signal is Smart Fusion (SF) [4]. It fuses three respiratory-induced

variations (RIIV, RIAV, RIFV) by calculating their mean and discards RR estimates that are labelled as artefact or where the standard deviation between the three estimates is larger than 4 breaths/min.

Research for improving the accuracy of RR estimation algorithms has primarily focused on discarding poor quality data points and did put less efforts in retaining or reconstructing information. For example, the SF discards a high percentage of estimates in order to retain only accurate RR information [4]. However, even if the PPG signal is shown to be of good quality, the RR is not computed due to the disagreements between different RR estimates. Pimentel et al. show that on average, 36% of data is eliminated in SF due to such disagreements [7]. Another estimation approach is to apply Kalman filters to each individual estimate leading to smoother time series for RR estimation [8]. However, this approach dampens rapid changes in RR and therefore reduces the detection of abnormal RR events that would be clinically relevant.

Abnormal event detection relies on robust and continuous RR measurements [9]. Robust measurements are defined by high accuracy of the estimation and low missing data, i.e. high retention rates after the fusion. It is crucial to retain as much continuous information as possible because data gaps might lead to missed signs of abnormal events and cause failure in recognizing patient deterioration [2], [10].

We present a novel fusion approach to extract RR from the PPG signal. Our aim was to develop an algorithm that 1) provides robust RR estimations independently of the quality of individual estimates, and 2) can be simultaneously optimized for clinical required accuracy and high retention rates. We characterized this algorithm against a benchmark dataset and the established SF algorithm.

II. METHODS

We proposed an algorithm that integrates covariance intersection fusion (CIF) as a central element to join the RRs extracted from the available PPG features (Fig. 1): 1) In a preprocessing step, the algorithm extracted the five RR induced variations from the PPG signal (RIAV, RIIV, RIFV, RISV, RIWV). 2) The dominant frequency for each variation was extracted from the maximum spectral power peak obtained by the FFT and a subtracted fitted power law function. 3) The quality of each estimation was determined by considering the noise index (NI) of each variation and an estimation was eliminated when this NI was too low. 4) Finally, the remaining RR estimates were fused into a final RR_{fusion} using the CIF method.

The work was supported by the Swiss National Science Foundation SNSF (150640).

Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland (email: jia.zhang@hest.ethz.ch, walter.karlen@ieee.org)

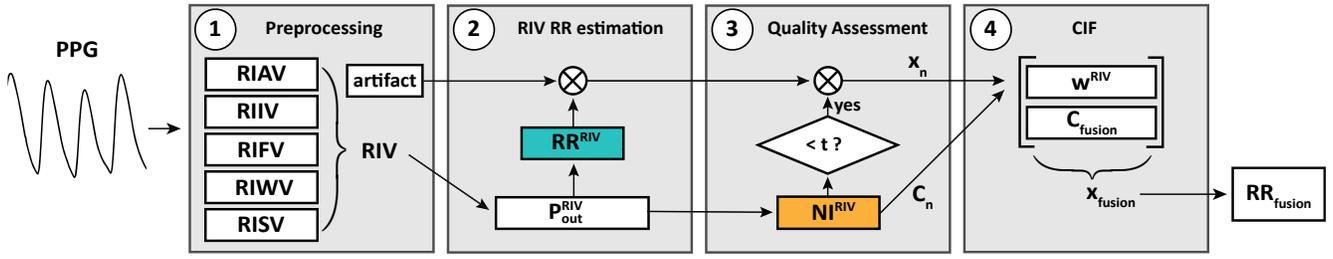


Fig. 1: RR covariance intersection fusion (CIF): The PPG waveform was preprocessed and five different respiratory-induced variations (RIV) and artifacts were extracted (1). Spectral analysis was conducted for each RIV to subsequently estimate the respiratory rate of each variation (RR^{RIV} , 2) and a noise index (NI^{RIV} , 3). The quality of each estimation was assessed by comparing NI to a threshold t . CIF fused all good quality RR estimates x_n using the NI derived covariance C_n (4) to compute a fused RR estimation (RR_{fusion}).

A. Preprocessing

We adopted the preprocessing steps from the SF approach that included filtering, pulse segmentation and artifact detection [4]. The RIAV, RIIV and RIFV were extracted using the same methods as described in [4]. The RIWV and RISV were extracted using the algorithms presented in [5] and [6]. Each RR induced variation (RIV) signal was re-sampled to 5 Hz.

B. Respiratory induced variation RR estimation

Analogous to the procedure in SF algorithm, when no artifact was detected we used the FFT to calculate the maximum spectral component within the RR band. The FFT window was 32 s with a shift of 2 s. We adapted the method by additionally introducing a power law fitting to reduce background noise [11]. First, we fitted a linear function into the log-log scale of the spectrum in the range of 2 to 4 breaths/min and 65 to 100 breaths/min, excluding the range of interest between 4 to 65 breaths/min such as

$$\log(P_{fit}) = a \times \log(f) + k, \quad (1)$$

$$P_{fit} = \exp^k \times f^a, \quad (2)$$

where P_{fit} was the fitted power spectrum, f was the frequency, a was the slope of the fitting line and k was the intersection. Very low (<2 breaths/min) and very high RR (>100 breaths/min) were not considered for the fitting because of possible ambiguous artefacts in these RR ranges. We then subtracted the fitted values from the power spectrum P such as

$$P_{out} = P - P_{fit}. \quad (3)$$

We detected the maximum peak $P_{out_{peak}}$ of the subtracted spectrum P_{out} within the RR range of interest ($j \in 4 - 65$ breaths/min) and its corresponding RR to be the estimate for each RR variation.

C. Estimate quality assessment

To evaluate the quality of each estimate, we calculated the NI such as

$$NI = \frac{P_{out_{peak}}}{\sum_j P_{out_j}}. \quad (4)$$

The NI described how discriminative the primary detected RR was from all other possible RR with a range from 0 to 1. If the NI for a RR estimate was smaller than a predefined threshold t , that estimate was not forwarded to the CIF.

D. Covariance intersection and RR fusion

CIF is an algorithm for fusing estimates with unknown correlation that origins from the field of control and estimation [12]. It considers the error correlation between the sources of estimates, which can achieve robustness and consistency in data fusion [12]. CIF takes the convex combination of the mean and covariance of the estimates expressed in information space. For an arbitrary number $n > 2$ of estimates (x_1, x_2, \dots, x_n), which are corrupted by measurement noises and modelling errors, the estimates can be fused into a final estimation x_{fusion} . The intersection is the convex combination of the covariances C_1, \dots, C_n , such as

$$C_{fusion}^{-1} = \omega_1 C_1^{-1} + \dots + \omega_n C_n^{-1}, \quad (5)$$

$$C_{fusion}^{-1} x_{fusion} = \omega_1 C_1^{-1} x_1 + \dots + \omega_n C_n^{-1} x_n, \quad (6)$$

$$\sum_{i=1}^n \omega_i = 1. \quad (7)$$

where ω_n was the weight of the estimate. The CIF algorithm estimated RR from multiple different RIVs obtained from a single PPG signal with unknown correlations. Therefore, the uncertainty of each estimate described by the NI was used as covariance term for the CIF, such as

$$C_n = 1 - NI_n. \quad (8)$$

E. Computational efficiency

We designed the CIF as a one-dimensional function. The condition

$$\omega_1 C_1 = \omega_2 C_2 = \dots = \omega_n C_n \quad (9)$$

solved eq. 7 for ω_n using n equations. This kept the computational effort with n division or multiplications relatively low compared to the previous processing steps. The entire algorithm was implemented in Matlab (R2019b, MathWorks Inc, Natick, USA).

F. Dataset

We tested our algorithm on the TBME RR benchmark dataset from CapnoBase [13]. The benchmark dataset contained 42 8-minute PPG recordings from 29 pediatric and 13 adult subjects undergoing elective surgery with either spontaneous or controlled ventilation. For comparison, the manually annotated data from capnography served as reference RR.

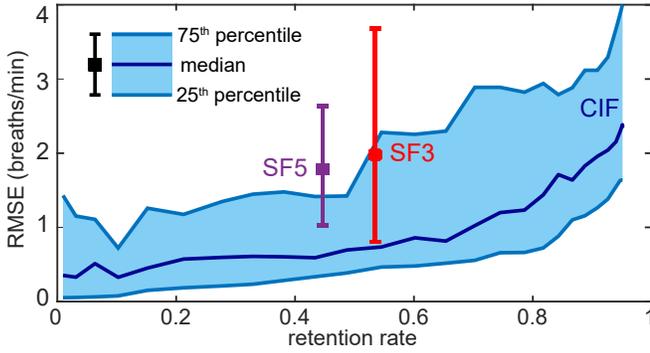


Fig. 2: Retention rate vs. RMSE of the proposed CIF obtained by varying the noise index threshold from 0 to 0.3 (blue shaded area). The RMSE increased with an increasing retention rate. The middle, lower and upper bound lines were median, 25th and 75th percentile of the RMSE for CIF, SF3, and SF5.

G. Evaluation

For a more objective comparison, we computed the SF using all the above mentioned five RIV estimates (SF5) in addition to the three original benchmark estimates (SF3) [4]. We evaluated the root mean squared error (RMSE) breath-by-breath across subjects for the fused RR against the reference RR from the benchmark dataset. We defined a retention rate as the ratio of the number of retained RR windows to the total possible number of RR windows within a recording. We varied the NI threshold t between 0 and 0.3 with an increment of 0.01 to depict the relationship between retention rate and RMSE.

To highlight the differences between the algorithms, we fixed t to 0.13 and calculated retention rate and RMSE for each subject. The results were depicted with boxplots. We conducted the Wilcoxon test with Bonferroni correction to compare the RMSE and retention rate distributions obtained by each algorithm. Finally, with computing Bland-Altman statistics and the Pearson's correlation coefficient r we investigated the agreement between the proposed CIF and the reference RR.

III. RESULTS

The RMSE for the CIF increased when the retention rate increased (Fig. 2). The SF5 showed lower retention rate than SF3 (Fig. 2). The proposed CIF showed lower RMSEs compared to the SF when keeping the retention rate fixed. The median RMSE over all 42 subjects was 1.4 breaths/min for the CIF compared to 1.8 breaths/min for the SF5 and 2.0 breaths/min for the SF3 (Fig. 3a). The CIF significantly increased the overall retention rate distribution from 0.46 (SF5) and 0.58 (SF3) to 0.90 ($p < 0.001$, Fig. 3b). Compared to the reference RR, r was 0.94 (Fig. 4a), the bias was 0.3 breaths/min and the limits of agreement were -4.6 and 5.2 breaths/min (Fig. 4b).

IV. DISCUSSION

We developed and evaluated a novel fusion algorithm to estimate RR from PPG signal based on five RR induced variations. The algorithm showed improved retention rates

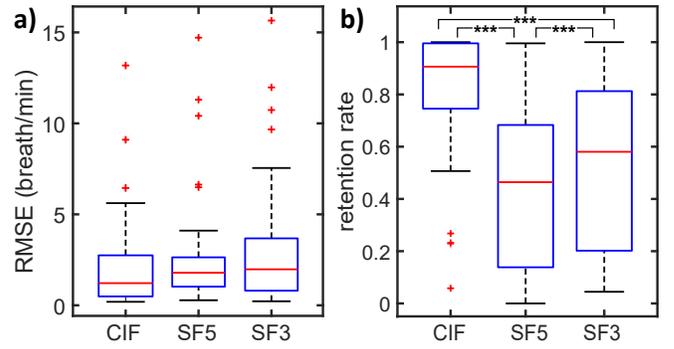


Fig. 3: Capnabase Benchmark comparison of CIF, SF with five estimates (SF5) and three estimates (SF3) for: a) RMSE and b) retention rate across subjects. The central, bottom and top edge of the box indicated the median, 25th and 75th percentiles. The outliers were depicted as a cross. The 3 asterisks indicated the significance level of the distribution difference (***) $p < 0.001$.

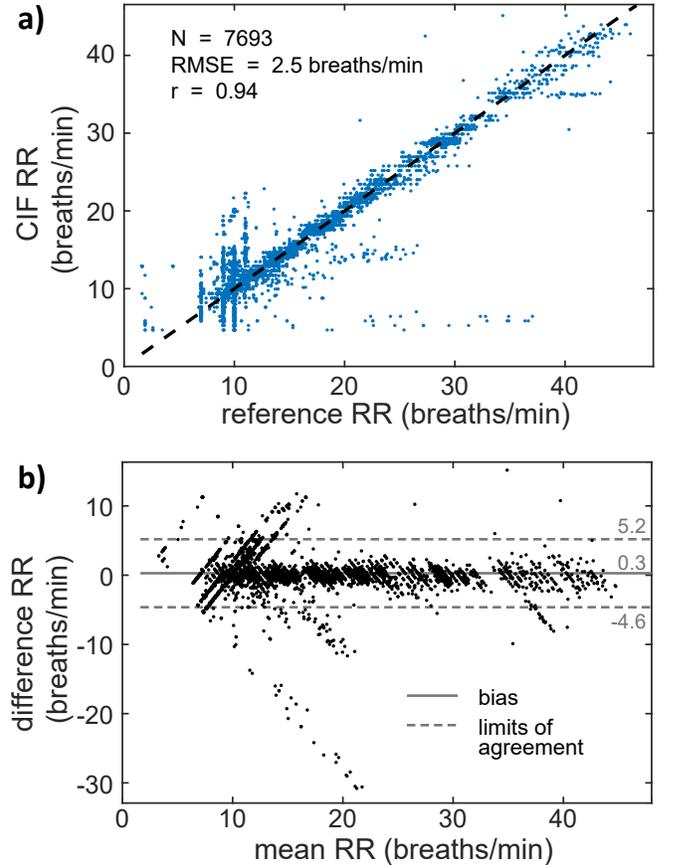


Fig. 4: Agreement between the CIF fused and the reference RR on the CapnoBase benchmark. a) scatter plot; b) Bland-Altman plot with the middle line representing the mean error (bias) and the two dash lines representing the limits of agreement.

and comparable high accuracy to the benchmark algorithm. The CIF based algorithm was adaptive and could fuse an arbitrary number of RR estimates independently of missing estimates that were eliminated during the quality control process. A simple parameter enabled the tuning of the performance.

The NI threshold offered an effective way to tune the

algorithm and optimize the RMSE and retention rate. There was an unavoidable trade-off between RMSE and retention rate. This method would enable a transparent way to align an application dependent task with a desired performance for the algorithm.

We evaluated the accuracy and retention rate of SF by fusing all five estimates and the three estimates used in [4]. Fusing all five estimates showed decreased median RMSE, but also decreased retention rate. This indicated that the two additional variations RIWV and RISV added value to the fusion. However, more estimates also introduced a higher degree of uncertainty that lead to less estimates. This indicated that considering only the agreement among estimates was not sufficient to extract a robust RR. More estimates were leading to higher likelihoods for the presence of a disagreement, leading to a lower retention rate. Under the consideration that all the RR estimates were extracted from a single PPG signal, causal dependencies among the estimates might have existed. Our algorithm did not treat the estimates independently, but considered the potential correlation and redundancy among estimates to reduce the fusion error. Therefore, our proposed method was more robust in terms of accuracy and retention rate compared to the SF algorithms.

The proposed CIF algorithm was computationally efficient. The average processing time for 30 RR estimations was 0.49 s on a single core CPU 2.60 GHz PC. Other approaches that focused on optimizing retention rate are computationally much more costly. For example, multiple steps of time-frequency dependencies (FFT, auto-correlation and auto-regression) need to be computed in [9]. For embedded applications power efficiency is key. Therefore, our approach offers a clear advantage for real-time assessments on wearables with limited computational power.

The CIF algorithm was evaluated only on one benchmark dataset that was limited in size and population. Evaluation on larger datasets including data that was acquired with wearable devices will be needed to better understand the performance of the CIF with altered sources of noise. For implementation, we recommend a prospective study with extensive validation experiments.

RR is one of the important vital signs that provides objective and convincing medical evidence to describe a patient's health status, yet it is often neglected [1], [2]. When continuous monitoring of RR estimation is essential for identifying abnormal events, low RR retention rates might lead to events missed. By simply considering the agreement among estimates, the current benchmark fusion method can easily fail when individual respiratory-induced variations are not prominent. This can be seen in specific diseases or populations, such as in critically ill or elderly where the absence of modulation from the autonomic system decreases or cancels the RIFV [14]. The main contribution of our proposed algorithm is that it can estimate the RR independently of the number of RIVs available. The decision to take a variation into consideration was taken at each fusion independently, making the system adaptive to changes in the input conditions.

For remote assessment applications, where medical experts are not available to operate the devices and interpret the measured signal, it is essential that the collected data are of high integrity [15]. In this work, we focused on improving two aspects of data integrity, i.e. accuracy (performance) and retention rate (completeness), and provided a tuning mechanism to optimize between the two. This approach is of high interest for robust RR estimation in wearable devices that measure PPG with optical sensors.

REFERENCES

- [1] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: The neglected vital sign," *Medical Journal of Australia*, vol. 188, no. 11, pp. 657–659, 2008.
- [2] W. Q. Mok, W. Wang, and S. Y. Liaw, "Vital signs monitoring to detect patient deterioration: An integrative literature review," *International Journal of Nursing Practice*, vol. 21, no. S2, pp. 91–98, 2015.
- [3] W. Karlen, H. Gan, M. Chiu, D. Dunsmuir, G. Zhou, G. A. Dumont, and J. M. Ansermino, "Improving the Accuracy and Efficiency of Respiratory Rate Measurements in Children Using Mobile Devices," *PLoS One*, vol. 9, no. 6, p. e99266, 2014.
- [4] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multi-parameter respiratory rate estimation from the photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013.
- [5] J. Lázaro, E. Gil, R. Bailón, A. Mincholé, and P. Laguna, "Deriving respiration from photoplethysmographic pulse width," *Medical and Biological Engineering and Computing*, vol. 51, no. 1-2, pp. 233–242, 2013.
- [6] P. S. Addison, "Slope Transit Time (STT): A Pulse Transit Time Proxy requiring only a Single Signal Fiducial Point," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2441–2444, 2016.
- [7] M. A. Pimentel, A. E. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Towards a Robust Estimation of Respiratory Rate from Pulse Oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1194–1923, 2017.
- [8] S. Nemati, A. Malhotra, and G. Clifford, "Data Fusion for Improved Respiration Rate Estimation," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 926305, 2010.
- [9] D. Birrenkott, M. A. Pimentel, P. J. Watkinson, and D. A. Clifton, "A Robust Fusion Model for Estimating Respiratory Rate from Photoplethysmography and Electrocardiography," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 2033–2041, 2018.
- [10] M. Prgomet, M. Cardona-Morrell, M. Nicholson, R. Lake, J. Long, J. Westbrook, J. Braithwaite, and K. Hillman, "Vital signs monitoring on general wards: Clinical staff perceptions of current practices and the planned introduction of continuous monitoring technology," *International Journal for Quality in Health Care*, vol. 28, no. 4, pp. 515–521, 2016.
- [11] J. M. Gottselig, C. L. Bassetti, and P. Achermann, "Power and coherence of sleep spindle frequency activity following hemispheric stroke," *Brain*, vol. 125, no. 2, pp. 373–383, 2002.
- [12] S. Julier and J. K. Uhlmann, *Handbook of Multisensor Data Fusion, Theory and Practice, Second Edition*. Boca Raton: CRC Press, 2009, ch. General decentralized data fusion with covariance intersection (CI).
- [13] W. Karlen, M. Turner, E. Cooke, G. A. Dumont, and J. M. Ansermino, "CapnoBase: Signal database and tools to collect, share and annotate respiratory signals," *Annual Meeting of the Society for Technology in Anesthesia (STA)*, p. 25, 2010.
- [14] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: a review of relevant physiology," *Journal of Medical Engineering & Technology*, vol. 36, pp. 1–7, 2012.
- [15] J. Zhang, L. Tüshaus, N. Nuño Martínez, M. Moreo, H. Verastegui, S. M. Hartinger, D. Mäusezahl, and W. Karlen, "Data Integrity Based Methodology and Checklist for Identifying Implementation Risks of Physiological Sensing in Mobile Health Projects: Quantitative and Qualitative Analysis," *JMIR mHealth and uHealth*, vol. 6, no. 12, p. e11896, 2018.