SAPPHIRE: STOCHASTICALLY ACQUIRED PHOTOPLETHYSMOGRAM FOR HEART RATE INFERENCE IN REALISTIC ENVIRONMENTS

B. Chwyl, A. G. Chung, R. Amelard, J. Deglint, D. A. Clausi, A. Wong

University of Waterloo
Systems Design Engineering
200 University Ave W, Waterloo, ON, Canada
{bchwyl, agchung, ramelard, jdeglint, dclausi, a28wong}@uwaterloo.ca

ABSTRACT

A novel method, **S**tochastically **A**cquired **P**hoto**p**lethysmogram for **H**eart rate **I**nference in **R**ealistic **E**nvironments (SAPPHIRE), is proposed for robust remote heart rate measurement through broadband video. A set of stochastically sampled points from the cheek region is tracked and used to construct corresponding time series observations via skin erythema transforms. From these observations, a photoplethysmogram (PPG) waveform is estimated via Bayesian minimization, with the required posterior probability inferred using a Monte Carlo approach. To mitigate the effects of noise, the contribution of each observation is weighted based on the observation's likelihood to contain relevant data. A bandpass filter is applied to the estimated PPG waveform to omit implausible heart rate frequencies, and the heart rate is estimated through frequency domain analysis. Experimental results acquired from a set of thirty videos indicate significantly improved performance in comparison to state-of-the-art methods.

***Index Terms*—** Photoplethysmography imaging, PPG, heart rate estimation, skin erythema, non-contact

1. INTRODUCTION

Heart rate is an important physiological indicator and is useful for diagnosing important health conditions such as cardiovascular disease [1], sleep apnea [2], and cognitive state [3]. Electrocardiograms (ECGs) [4] are the current gold standard for heart rate measurement; however, ECGs require adhesive patches and straps, resulting in patient discomfort given extended periods of use. In addition, the relatively high cost of ECG machines makes them inaccessible and infeasible for widespread personal use. Pulse oximeters [5] are a low cost alternative which construct a photoplethysmogram (PPG) waveform, a representation of the change in blood volume over time at a particular location, by measuring the light transmitted through an extremity (e.g., finger, ear lobe) over time. Though non-invasive, pulse oximeters rely on contact measurements, limiting its use to situations where constant physical contact is feasible.

Recently, there has been a push towards remote heart rate estimation via photoplethysmography imaging (PPGI) [6], the construction of a PPG waveform through video. This is achieved in a similar manner as contact PPG techniques, however, rather than measuring the light transmitted through an extremity, the light reflected from a skin region is measured. PPGI systems are more sanitary, efficient, and less obtrusive than contact PPG systems. In addition, PPGI systems facilitate newer applications such as telemedicine, affective computing, and monitoring a subject in motion. Methods exist which make use of active illumination [7, 8, 9] or filtered light [10, 11] to accurately construct PPG waveforms; however, these methods typically require custom hardware or involved configurations. Due to the ubiquity of digital cameras, measuring the reflectance of ambient light with broadband red, green, and blue channel (RGB) cameras offers a readily available means of acquiring a PPG waveform, greatly increasing general accessibility and decreasing cost.

Broadband PPGI methods [12, 13, 14] which use consumer-level cameras to measure changes in reflected ambient light caused by varying hemoglobin concentration throughout a video sequence have been explored. As hemoglobin concentration is indicative of blood, these changes in reflected light can be used to construct a PPG waveform. From analyzing the time between peaks or the frequency content of the constructed PPG waveform, an individual's heart rate can be estimated.

Poh *et al.* [12, 13] proposed two methods for remote heart rate estimation from broadband video. Both methods estimate heart rate by first taking the average red, green, and blue pixel values within a window around an individual's face to produce raw RGB signals. Independent component analysis (ICA) is performed to decompose the raw RGB signals into three independent source signals for analysis. Filters are applied to both the raw signals and the source signals to mitigate noise. In their initial method [12], the second source signal is selected and analyzed in the frequency domain to estimate heart rate. In their more recent work [13], a custom

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Research and Innovation, and the Canada Research Chairs Program.

peak detection algorithm is applied to the source signal with the maximum magnitude in the frequency domain.

While both methods proposed by Poh *et al.* reported good results, the first method [12] is subject to error due to the random order in which the ICA algorithm returns the source signals. Additionally, error can be introduced by irrelevant data (i.e., hair, mouth) contained in the large sample region. Their second method [13] addressed the issue of the random source signal order by automatically selecting the signal with the maximum peak in the frequency domain; however, the problem of capturing irrelevant data in the sample region remains a possible source of error.

Another method was proposed by Li *et al.* [14] in which a PPG waveform is constructed by averaging the green channel values of pixels within a region defined by facial landmarks. The acquired signal is filtered and de-trended before the frequency content is analyzed to estimate the heart rate. The average green channel of the scene's background is also subtracted from the PPG waveform to compensate for the effects of ambient illumination changes. However, the de-trending and filtering aspects of this method risk degrading features of the PPG waveform useful for heart rate estimation.

Although current PPGI methods are able to extract an individual's heart rate from video, the results are often inconsistent due to the encapsulation of irrelevant data, lack of consistency in selecting relevant signals, or over-filtering. Furthermore, scenes with uneven illumination, dynamic backgrounds, partial occlusion, or motion can cause inaccuracies in state-of-the-art methods. For these reasons, a robust and consistent broadband PPGI method capable of constructing a valid PPG waveform in the presence of noise is required. As such, we propose SAPPHIRE: **S**tochastically **A**cquired **P**hotop**l**ethysmogram for **H**eart rate **I**nference in **R**ealistic **E**nvironments, a method in which a statistical framework is used to produce a PPG waveform estimate resilient to noise.

2. METHODS

SAPPHIRE relies on the accurate estimation of a PPG waveform. We first estimate a PPG, $\hat{\phi}(t)$, via a Bayesian minimization approach, using a set of skin erythema signals as observations measured from points stochastically sampled from the cheek region and tracked throughout the video. The posterior probability required for the Bayesian estimation is estimated through an importance-weighted Monte Carlo sampling approach, in which observations likely to yield valid PPG data are predominant. A Fourier transform is applied to the estimated PPG waveform and the frequency bin corresponding to the maximum peak within an operational band ($[f_l - f_h]Hz$), is selected as the heart rate frequency, HR_{Hz}. It is then converted into the standard unit of beats per minute (bpm) by multiplying by 60 to produce the final estimate, HR_{bpm}. A flow diagram outlining the SAPPHIRE algorithm can be seen in Fig. 1.

2.1. Sample Points

Determining appropriate sample locations is an important aspect of constructing a clean PPG waveform. Due to the low facial skin thickness [15], flatness of the area, and lower chance of occlusion from facial hair, hats, scarves, etc., an initial set of points is stochastically sampled from the cheek region. This region is determined in relation to the face and eyes, registered using a cascade object detection algorithm [16] on the initial video frame. Once located, these points are tracked throughout the video using the Kanade-Lucas-Tomasi (KLT) tracking algorithm [17].

2.2. Skin Erythema

Because of skin erythema's high correspondance to hemoglobin concentration [18], its fluctuation offers a biologically motivated method for remote heart rate estimation [19]. Though many skin erythema models exist [20, 21, 22], we use the model proposed by Yamamoto *et al.* [23] due to the broadband nature of our acquired signal. The skin erythema transform, $\psi(t)$, for a single point, x, can be expressed as

$$\psi(t) = \log_{10} \frac{1}{\bar{x}_g(t)} - \log_{10} \frac{1}{\bar{x}_r(t)}, \tag{1}$$

where $\bar{x}_g(t)$ and $\bar{x}_r(t)$ represent the average values of the green and red channels, respectively, within a window of size $\aleph \times \aleph$ surrounding x. The spatial average is taken in order to mitigate the effects of point noise and tracking error.

2.3. Bayesian Minimization

To produce an accurate heart rate estimation, it is important to reliably construct an estimation of a PPG waveform, $\phi(t)$. By using a statistical framework to consider the contributions of numerous observations, an accurate estimate of $\phi(t)$ can be produced. This is achieved via a Bayesian minimization approach, where a set of skin erythema time series is used as observations. Each skin erythema time series is constructed by applying the skin erythema transform, $\psi(t)$, to a specific point at each time step as it is tracked throughout the video. The Bayesian minimization is formulated as

$$\hat{\phi(t)} = \arg\min_{\hat{\phi(t)}} (E((\hat{\phi(t)} - \phi(t))^2 | X)), \tag{2}$$

where $E(.)$ denotes the expectation, and X represents the set of accepted observations. By taking the derivative and setting the equation equal to zero, this formula can be solved and simplified to yield

$$\hat{\phi(t)} = \int \phi(t) p(\phi(t)|X) d\phi(t). \tag{3}$$

It can be seen that the posterior probability, $p(\phi(t)|X)d\phi(t)$, is required to solve the minimization, however, it is difficult to obtain analytically. For this reason, an importance weighted Monte Carlo sampling approach is used [24, 25].

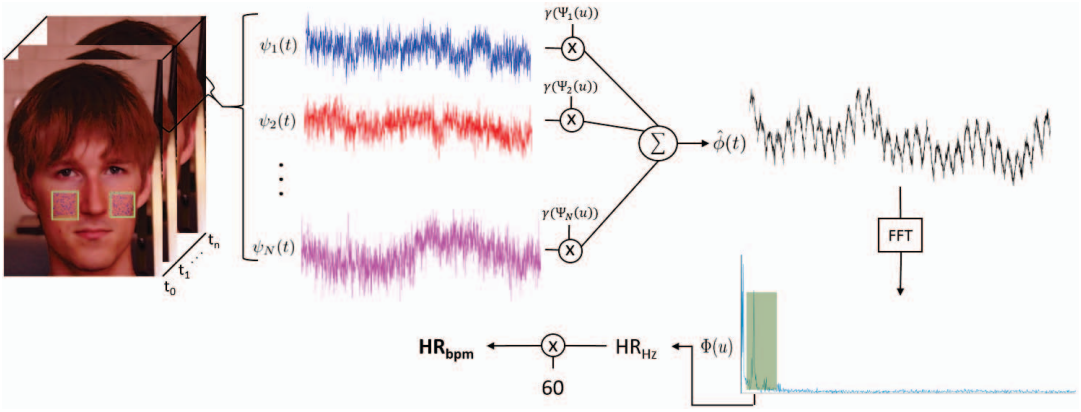

Fig. 1. Flow chart of the SAPPHIRE algorithm. Bayesian minimization is used to estimate a PPG wave form, $\hat{\phi}(t)$, from multiple skin erythema time series ($\psi(t)$). The contributions for each signal are weighted based on the function $\gamma(\Psi_i(u))$, which considers the Fourier transform of $\psi(t)$, $\Psi(u)$. From the frequency content of $\hat{\phi}(t)$, $\hat{\Phi}(u)$, a heart rate can be inferred within an operational band (indicated in green).

2.4. Posterior Probability Estimation

In this approach, the set of accepted observations, X, must first be established. The likelihood that an observation, x_i, is added to the set X is determined as

$$\gamma(\Psi_i(u)) = \exp\left(-\alpha \frac{\max(\Psi_i(u))}{NF}\right) \quad \text{subject to} \quad f_l \leq u \leq f_h, \tag{4}$$

where $\Psi(u)$ is the Fourier transform of $\psi_i(t)$, NF is the noise floor of $\Psi(u)$, and α is an empirically determined threshold. Since signals corrupted by noise provide little relevant data, a distinct response in the frequency domain is unlikely and such signals will have a low probability of being added to X. With the set X established, the posterior probability can be estimated as a weighted histogram:

$$\hat{p}(\phi(t)|X) = \frac{\sum_{i=0}^{N} \gamma(\Psi_i(u))\psi_i(t)}{\sum_{i=0}^{N} \gamma(\Psi_i(u))}, \tag{5}$$

where N is the number of points in the set X. The acceptance probability is also used as a weighting, designed to ensure the contributions of cleaner signals is greater than those contaminated by noise. With the posterior probability obtained, the estimated PPG waveform, $\hat{\phi}(t)$, can be constructed via Eq. 3.

2.5. Frequency Domain Analysis

To extract a heart rate estimate, the estimated PPG waveform is transformed into the frequency domain ($\hat{\phi}(t) \xrightarrow{\mathfrak{F}} \hat{\Phi}(t)$). The heart rate is then estimated as

$$HR_{Hz} = \arg\max_u |\hat{\Phi}(u)| \quad \text{subject to} \quad f_l \leq u \leq f_h. \tag{6}$$

Finally, HR_{Hz} is multiplied by 60 to obtain HR_{bpm}, a heart rate estimate in the standard unit of measurement.

3. EXPERIMENTS

3.1. Experimental Setup

To verify the validity of our method, six videos were recorded for each of five subjects with a Point Grey Chameleon3 camera at 80 frames per second in natural ambient lighting conditions. The videos were divided into two categories for each subject: four stationary videos, where the subject was sitting still, and two gradual motion videos, where the subject was free to move and speak to mimic realistic testing scenarios.

Ground truth data was acquired using a finger pulse oximeter (Embedded Lab's Easy Pulse 1.1) which was wired to illuminate a red LED with each heart beat. The red LED was positioned in view of the camera behind the subject as to not inadvertently illuminate the subject's face. The red LED was used to produce a time series, and the ground truth heart rate was determined by the frequency corresponding to the maximum peak in the frequency domain of the LED signal. An example frame from the recorded data set demonstrating the configuration can be seen in Fig. 2.

To compare against state-of-the-art methods, both methods proposed by Poh *et al.* [12, 13] as well as the method proposed by Li *et al.* [14] were implemented as closely as possible. Since the underlying implementation of the custom peak detector used by Poh *et al.* [13] is not described, we instead used frequency domain analysis to estimate the heart rate in order to remain consistent with the other methods. In addition, parameter values were acquired from the corresponding papers where mentioned and empirically determined otherwise. In our method, the parameter α was set to 2 and the window size, $\aleph$, was set to 5 as these values were empirically determined to produce the best results. The operational band was chosen as $f_l = 0.7Hz$ and $f_h = 4.0Hz$ as it provides a wide range of heart rate values ($[42 - 240]bpm$) and encompasses average resting heart rates [26].

Table 1. Results of state-of-the-art comparison subdivided into three categories: stationary, gradual motion, and all videos. The mean error (M_e), standard deviation (σ_e), absolute mean error ($|M_e|$) and root mean squared error ($RMSE$) are tabulated with the best results for $|M_e|$ and $RMSE$ indicated in boldface. All results are shown in terms of beats per minute (bpm).

	Stationary			Gradual Motion			Both		
	M_e (σ_e)	$\|M_e\|$	$RMSE$	M_e (σ_e)	$\|M_e\|$	$RMSE$	M_e (σ_e)	$\|M_e\|$	$RMSE$
SAPPHIRE	-5.78 (15.77)	**7.79**	**16.80**	-11.28 (18.66)	**12.47**	**21.80**	-7.61 (16.99)	**9.36**	**18.62**
Poh 2010 [12]	-12.71 (17.19)	14.14	21.38	-10.65 (28.48)	22.81	30.41	-12.02 (21.64)	17.03	24.76
Poh 2011 [13]	29.88 (24.04)	30.77	38.35	31.74 (24.39)	32.07	40.03	30.50 (24.17)	31.20	38.92
Li 2014 [14]	20.25 (21.12)	21.16	29.26	15.19 (18.24)	18.35	23.74	18.57 (20.35)	20.22	27.54

Fig. 2. Example frame from our recorded data set. Ambient light was used to illuminate the face while ground truth data was acquired from a red LED, which turned on with each heart beat, placed behind the subject.

3.2. Experimental Results

The mean error (M_e), standard deviation (σ_e), absolute mean error ($|M_e|$), and root mean squared error ($RMSE$) were calculated in beats per minute (bpm) for each of the methods using our data set (Table 1). It can be seen that in most cases, our method produces mean error closest to zero with the lowest standard deviation, implying more consistently accurate results. In addition, SAPPHIRE achieves the lowest absolute error and root mean squared error.

The methods proposed by Poh *et al.* (2011) [13] and Li *et al.* [14] produce a positive mean error, indicating erroneously high heart rate estimations. This is likely due to the de-trending and filtering procedures used to suppress low frequency content, resulting in suppressed relevant frequency content and leading to correct peaks in the frequency domain being overshadowed by high frequency content. Conversely, Poh *et al.* (2010) [12] often produces estimates lower than expected. This is likely the result of analyzing an incorrect signal after source signal separation. With no distinct peak in the range of feasible heart rates, the inherently high magnitude of low frequency content is incorrectly selected.

A potential source of error for SAPPHIRE is the incorrect registration of low frequency content as the heart rate frequency. This is likely caused by the quasi-periodic nature of the PPG waveform as well as minor fluctuations in heart rate throughout some of the test videos. This results in a frequency response across multiple frequency bins, leading to the inherently high response of low frequency content dominating the correct peaks.

A t-test was performed on the absolute error distributions produced by Poh *et al.* (2010) [12], Poh *et al.* (2011) [13], and Li *et al.* [14], across all videos (stationary and gradual motion). This yielded p-values of 0.0065, 0.0003, and 0.0182, respectively, indicating statistical significance within a 95% confidence interval.

4. CONCLUSIONS

A novel statistical framework for remote heart rate estimation from broadband video using stochastic Bayesian estimation was proposed. Preliminary results using a set of 30 videos under ambient lighting conditions show 7.67 bpm and 6.14 bpm decreases in absolute mean error and root mean squared error, respectively, indicating a significant improvement in accuracy over current state-of-the-art methods.

Future work will include the exploration of temporal PPG waveform analysis to allow for the estimation of inter beat intervals, and the use of short time Fourier transforms as a means of estimating heart rate variability. In addition, further testing against a greater number of videos will be performed to ensure robustness.

5. REFERENCES

[1] Javaid Nauman, Imre Janszky, Lars J Vatten, and Ulrik Wisløff, "Temporal changes in resting heart rate and deaths from ischemic heart disease," *JAMA*, vol. 306, no. 23, pp. 2579–2587, 2011.

[2] Martica Hall, Raymond Vasko, Daniel Buysse, Hernando Ombao, Qingxia Chen, J David Cashmere, David Kupfer, and Julian F Thayer, "Acute stress affects heart rate variability during sleep," *Psychosomatic Medicine*, vol. 66, no. 1, pp. 56–62, 2004.

[3] Daniel McDuff, Sarah Gontarek, and Rosalind Picard, "Remote measurement of cognitive stress via heart rate variability," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 2957–2960.

[4] Louis Nelson Katz and Alfred Pick, *Clinical electrocardiography*, Lea & Febiger, 1956.

[5] Takuo Aoyagi, "Pulse oximetry: its invention, theory, and future," *Journal of Anesthesia*, vol. 17, no. 4, pp. 259–266, 2003.

[6] Markus Huelsbusch and Vladimir Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using ppgi," in *Medical Imaging 2002*. International Society for Optics and Photonics, 2002, pp. 110–117.

[7] FP Wieringa, Frits Mastik, and AFW Van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: a first step toward spo2 camera technology," *Annals of Biomedical Engineering*, vol. 33, no. 8, pp. 1034–1041, 2005.

[8] Giovanni Cennini, Jeremie Arguel, Kaan Akşit, and Arno van Leest, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Optics Express*, vol. 18, no. 5, pp. 4867–4875, 2010.

[9] Yu Sun, Sijung Hu, Vicente Azorin-Peris, Stephen Greenwald, Jonathon Chambers, and Yisheng Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of Biomedical Optics*, vol. 16, no. 7, pp. 077010–077010, 2011.

[10] Mark van Gastel, Sander Stuijk, and Gerard de Haan, "Motion robust remote-ppg in infrared," *Biomedical Engineering, IEEE Transactions on*, vol. 62, no. 5, pp. 1425–1433, 2015.

[11] Lingqin Kong, Yuejin Zhao, Liquan Dong, Yiyun Jian, Xiaoli Jin, Bing Li, Yun Feng, Ming Liu, Xiaohua Liu, and Hong Wu, "Non-contact detection of oxygen saturation based on visible light imaging device using ambient light," *Optics Express*, vol. 21, no. 15, pp. 17464–17471, 2013.

[12] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[13] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.

[14] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 4264–4271.

[15] Richard Y Ha, Kimihiro Nojima, William P Adams Jr, and Spencer A Brown, "Analysis of facial skin thickness: defining the relative thickness index," 2005.

[16] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 1, pp. I–511.

[17] Carlo Tomasi and Takeo Kanade, *Detection and tracking of point features*, School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

[18] Hao Gong and Michel Desvignes, "Hemoglobin and melanin quantification on skin images," in *Image Analysis and Recognition*, pp. 198–205. Springer, 2012.

[19] Audrey Chung, Xiao Yu Wang, Robert Amelard, Christian Scharfenberger, Joanne Leong, Jan Kulinski, Alexander Wong, and David A Clausi, "High-resolution motion-compensated imaging photoplethysmography for remote heart rate monitoring," in *SPIE BiOS*. International Society for Optics and Photonics, 2015, pp. 93160A–93160A.

[20] JB Dawson, DJ Barker, DJ Ellis, JA Cotterill, E Grassam, GW Fisher, and JW Feather, "A theoretical and experimental study of light absorption and scattering by in vivo skin," *Physics in Medicine and Biology*, vol. 25, no. 4, pp. 695, 1980.

[21] BL Diffey, RJ Oliver, and PM Farr, "A portable instrument for quantifying erythema induced by ultraviolet radiation," *British Journal of Dermatology*, vol. 111, no. 6, pp. 663–672, 1984.

[22] JW Feather, DJ Ellis, and G Leslie, "A portable reflectometer for the rapid quantification of cutaneous haemoglobin and melanin," *Physics in Medicine and Biology*, vol. 33, no. 6, pp. 711, 1988.

[23] Tadamasa Yamamoto, Hirotsugu Takiwaki, Seiji Arase, and Hiroshi Ohshima, "Derivation and clinical application of special imaging by means of digital cameras and image j freeware for quantification of erythema and pigmentation," *Skin Research and Technology*, vol. 14, no. 1, pp. 26–34, 2008.

[24] W Keith Hastings, "Monte Carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[25] Dorothy Lui, Amen Modhafar, Jeffrey Glaister, Alexander Wong, and Masoom A Haider, "Monte Carlo bias field correction in endorectal diffusion imaging," *Biomedical Engineering, IEEE Transactions on*, vol. 61, no. 2, pp. 368–380, 2014.

[26] "Target heart rates," http://www.heart.org, Accessed: 2016-01-25.