



Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

[Laurie, Jordan, Higgins, Niall, Peynot, Thierry, & Roberts, Jonathan](#)
(2020)

Dedicated exposure control for remote photoplethysmography.
IEEE Access, 8, pp. 116642-116652.

This file was downloaded from: <https://eprints.qut.edu.au/202877/>

© Consult author(s) regarding copyright matters

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution 4.0

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

<https://doi.org/10.1109/ACCESS.2020.3003548>

Received May 24, 2020, accepted June 4, 2020, date of publication June 25, 2020, date of current version July 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003548

Dedicated Exposure Control for Remote Photoplethysmography

JORDAN LAURIE¹, NIALL HIGGINS², THIERRY PEYNOT¹, (Member, IEEE),
AND JONATHAN ROBERTS¹, (Senior Member, IEEE)

¹School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, QLD 4000, Australia

²School of Nursing, Queensland University of Technology, Brisbane, QLD 4000, Australia

Corresponding author: Jordan Laurie (jlaur21@gmail.com)

ABSTRACT This paper aims to show that control of exposure time during video capture will improve the accuracy of remote photoplethysmography (rPPG). We propose a purpose specific exposure control algorithm for use in heart rate estimation via rPPG applicable for any controllable camera. Our novel algorithm works by selecting exposure that achieves maximum Signal-to-Noise Ratio (SNR) before distortion will occur. We performed experiments to test the accuracy of non-contact PPG extracted simultaneously from two identical cameras positioned together but with different exposure time controls. Our purpose specific algorithm in camera A controlled exposure time to maximise rPPG SNR ratio while camera B remained set at one of a range of values. Exposure time set by our novel algorithm out-performed camera B with a lower mean absolute error relative to a standard pulse oximeter. A significant improvement to heart rate estimation performance using a research camera can be made with specific control of exposure time. The improvements in performance demonstrated here are an important step in taking rPPG out of a lab environment and into less controlled circumstances such as clinical settings and emergency rescue scenarios.

INDEX TERMS Biomedical engineering, biomedical monitoring, computer vision.

I. INTRODUCTION

Vital signs assessment using cameras, known as remote photoplethysmography (rPPG), has shown significant promise in recent years [6], [8], [11]. One of the current practical limitations is the ability to perform under everyday working conditions. Performance is very strong under desirable circumstances. There has been little reporting on how performance changes or degrades as conditions become less desirable. This difficulty in less controlled environments is a key limitation in the clinical usage of rPPG. Possible clinical applications range from situations where non-contact vital signs assessment would be more conducive to patient care in a mental health facility where low stimulus is desired, through to emergency response for a fallen victim in difficult terrain where these observations could be obtained using a robot or drone fitted camera and then relayed to the rescue team for triage.

Photoplethysmography (PPG) uses reflected light to quantify cutaneous blood volume [13], [19], [22]. Light emitted

The associate editor coordinating the review of this manuscript and approving it for publication was Tomasz Trzcinski.

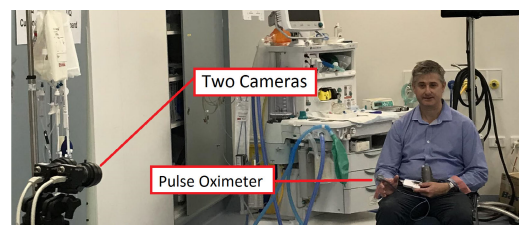


FIGURE 1. Controlling exposure will improve performance when encountering uncontrolled lighting and distances as shown in this clinical setting.

from a contact PPG device is constant and relatively homogeneous but the variability in natural lighting can be unpredictable in both intensity and duration. The challenge that rPPG methods face is to minimize noise at different levels of available light.

Little attention has been paid to camera parameters at the time of image capture. The camera parameters most relevant to rPPG extraction are those that control exposure: aperture size, gain and exposure time (also called shutter speed). Appropriate control of the duration of exposure time may improve the Signal-to-Noise Ratio (SNR) without causing

saturation distortion and produce video more suited to rPPG extraction. This will lead to reduced error in heart rate (HR) estimation. This is particularly true in challenging settings such as those involving poor lighting or greater distances.

Here we propose a system for direct purpose specific exposure control for the function of HR assessment via rPPG. We will review of the current state of the art for HR assessment that use standard RGB cameras. We will then direct our attention to direct control of the camera and its potential for performance improvement. We propose a method for exposure control that is purpose specific to HR assessment through rPPG. This system was then implemented together with a state of the art rPPG estimation algorithm. Experimental validation of our proposed method was conducted with two cameras: Camera A running both the aforementioned state of the art rPPG and our novel direct exposure control, and Camera B running the same rPPG algorithm only. The performance of both of these cameras for HR estimation was assessed relative to a standard pulse oximeter. Potential future improvements of our algorithm are discussed, including enabling our algorithm to make larger individual changes in exposure time and removing frame rate limitations.

II. BACKGROUND

A. HEART-RATE ESTIMATION WITH A CAMERA

There have been several studies into camera-based HR extraction with all methods following a similar core procedure: (i) capture video of exposed skin; (ii) select a Region-of-Interest (ROI) and use the mean pixel intensity from each frame to create a time series; (iii) extract a PPG from these time series; and finally (iv) estimate the fundamental frequency of the PPG - which is assumed to be HR.

1) CAPTURE VIDEO

The first step in estimating rPPG is to capture a video. There are two main methods for doing this. The most accepted method for rPPG extraction employs normal room lighting and standard RGB cameras [26]. This approach of using ambient lighting conditions continues to be used, in part due to it being readily accessible [6], [8], [11].

2) SELECT ROI TO CREATE TIME SERIES

Recent developments have automated ROI selection so that this function is integrated with the rest of the HR extraction procedure [1], [6], [8], [11]. These studies initially employed the Viola-Jones (VJ) face identification algorithm. The VJ algorithm creates a best-fit rectangle of the region [27]. While some studies have used the VJ algorithm to detect the face before every frame, this approach is not appropriate for real-time estimation due to the slow nature of this process. A more efficient approach has been to apply the Kanade-Lucas-Tomasi (KLT) feature tracker to subsequent frames once the face is identified [24].

3) EXTRACT PPG

When a set of complete time series have been obtained from the ROI means, an rPPG estimate is made. The intention is to

initially estimate an rPPG with the highest possible signal-to-noise ratio (SNR). Most calculations are based upon the time series obtained from each of the RGB colour channels and appear to produce more accurate rPPG estimates than single channel methods [5]–[7], [10].

There are two main approaches to extracting the PPG from the colour channel time series. In Blind Source Separation (BSS) based approaches the three colour channels are unmixed into three signals, one of which is presumed to be the PPG, using techniques such as Independent Component Analysis, Principle Component Analysis and Canonical Correlation Analysis [1], [10]. In chrominance based methods each colour channel is normalized (in small, overlapping time windows) and bandpass filtered. This process has achieved results comparable to BSS based methods [3]–[5], [8], [28]. De Haan et al extended this work by integrating the product of the sensitivity curves of all three colour channels with reflectance curves obtained from early studies on illumination of subcutaneous blood vessels to create a “signature” that was then used to mathematically weight the colour channels before combining them [4]. In addition to his approach to ROI selection, Feng refined the chrominance technique to produce an estimate of rPPG based upon the raw and filtered (HR range) colour channels [6]. Feng’s method has been chosen for rPPG extraction used in this paper.

4) ESTIMATE HEART RATE

Once the rPPG is produced, the HR itself can be estimated. The rate at which blood pumping cycles are completed by the heart is the fundamental frequency of the rPPG.

The most commonly used method of estimating the fundamental frequency of a uniform time-series is to identify the dominant peaks from a Fast Fourier Transform (FFT) within an identified window of time [4], [5], [25], [26]. The benefits of FFT are well understood but the main disadvantage with using this approach for rPPG is its lack of frequency resolution.

Current methods of rPPG extraction from video have demonstrated the ability to produce HR estimates from a standard RGB camera. Results are encouraging and point to potential future clinical usage. In the current literature, improvements in accuracy are achieved through video processing methods. There has been no investigation into improvement from dedicated capture of video for this purpose. Control of exposure is a potential source of improvement in rPPG SNR, and by extension a reduction in error in HR estimation.

B. EXPOSURE CONTROL

The most prominent aim in exposure control literature is to produce aesthetic frames for consumer electronics. Exposure of a video frame is the duration and amount of light that the image sensor receives and is measured in lux seconds. Overall exposure is controlled by three specific parameters: aperture size, gain and exposure time. Aperture size describes the diameter of the opening in the lens through which light

enters the camera. A larger aperture allows more light to hit the image sensor, increasing the exposure of a given frame. Changing the aperture size also changes the focus of the frame. However, aperture size is often not a controllable parameter as this is determined during the manufacture of the camera hardware. Gain is an analogue amplification applied to the image sensor output in the hardware of the camera, often referred to as “ISO”. The fact that gain is applied before quantization means that it is not equivalent to increasing brightness through digital processing. Increasing gain increases the exposure of the frame but does not increase the SNR.

Exposure time is the amount of time the open shutter of the camera exposes the image sensor to light. Exposure time is often software controllable. Since the image sensor is in effect integrating over this period, larger exposure times make each frame a less instantaneous sample and may lead to effects such as motion blur. The maximum exposure time is also constrained by the frame rate of the recording.

The problem of achieving a desired exposure automatically has been studied, partially driven by the need for commercial cameras to have an automatic exposure control mode whilst simultaneously achieving an aesthetically pleasing image. Measurement of exposure must be tied to some measure in the frame in the absence of an external metering device; these measures are routinely based on the distribution of pixel intensities and statistics of the distribution [9], [20]. Algorithms are then employed to attempt to adjust overall exposure in order to have this measure reach its desired value. We describe here only those uses of exposure control that do not use external metering devices.

The literature contains several approaches to choosing a measure of correctness for exposure of a frame. Popular modern methods use a histogram of pixel intensities to select a desired brightness level [20]. These methods may use the image as a whole, or concentrate on specific regions of the frame [12]. Once a target for exposure is chosen, the camera must then attempt to respond accordingly. What these methods have in common is an emphasis on exposure in the context of producing frames desirable to a human viewer and not the purpose of obtaining a better PPG.

Early work did examine the overall exposure in context of all parameters [15] and within the photography related literature changing exposure time is most popular choice from the three specific exposure parameters. A simple model of automatic exposure control (AEC) as the time the sensor integrates some constant light striking it implies linear scaling, for example:

$$\int_0^{aT_0} L dt = aT_0L \quad (1)$$

where L is the rate at which light is striking the image sensor, T_0 is the exposure time, a is a constant and t is time.

However, the slight non-linearity of the relationship in practice is often the primary obstacle to be overcome in trying to achieve a desired brightness level in the frame. If the

exact relationship between exposure time and pixel intensities is unknown, the problem of achieving a desired intensity becomes one of root finding on an unknown function. This problem is well studied mathematically and numeric approaches have shown success in reaching a desired exposure quickly and accurately [21], [23].

Studies aimed at achieving an exposure time for a specific purpose other than producing aesthetic images are lacking [16]. One example of purpose specific exposure control was described by Nuske. He intentionally applied different levels of exposure to subsequent image frames to achieve desired exposure in all parts of a frame. A selection was then made from a set of frames based on a specific object within the image to assist with guiding an autonomous vehicle [17], [18]. This idea of purpose specific control of exposure focused on a region is of particular interest in the case of rPPG as we are only interested in the exposure for one very specific part of the frame.

III. METHOD

It is necessary to understand the effect of noise on the ROI to understand our specific approach to exposure control. We will first examine the role of gain and the effect of quantization noise on the ROI mean. Then we will assess shutter speed and how long the image sensor is exposed to the light entering the camera. We will first examine the effect that exposure time has on random noise and on the ROI. Changing the aperture size also changes the frames depth of field. However, we will ignore techniques that control aperture size as this is not often controlled by software and is therefore less generally applicable to HR assessment using cheap, off the shelf cameras.

A. ENSEMBLE QUANTIZATION NOISE

Gain, describes the amplification that is applied to the image sensor’s analog output. It is important that this amplification is applied before quantization so it has the effect of maximizing signal-to-quantization-noise ratio (SQNR). Amplifying the frame digitally will not have this effect. The quantization noise affecting an ROI mean is more complicated than the quantization noise affecting a single sensor - as the ROI could be thought of as an ensemble of individual measurements affected by quantization. There will be a different level of noise reduction if gain is applied to the whole ROI. We assume each pixel in the ROI is affected by noise independently, not every possible source of noise acts independently, random electronic noise dominates in recording such as ours and can be reduced by increasing exposure time. We will explicitly consider how this changes the SNR when gain is increased. We assume all pixels in the ROI observe some true value u , that falls between quantization intervals α and $\alpha + \Delta$, although it is closer to α . With zero noise all pixels would quantize to α for a mean of α . If instead each pixel observes $u + n$, where $n \sim \mathcal{N}(0, \sigma)$, those pixels with higher noise will now be quantized to $\alpha + \Delta$. The ensemble mean will now be closer to u .

If the noise is powerful enough it may cause pixels to quantize to intervals further from u , such as $\alpha - \Delta$ or $\alpha + 2\Delta$. Let X be an interval, the probability that a pixel will quantize to interval X is:

$$p(X) = P\left(u + n < x + \frac{\Delta}{2}\right) - P\left(u + n < x - \frac{\Delta}{2}\right) \quad (2)$$

$$p(X) = P\left(n < x - u + \frac{\Delta}{2}\right) - P\left(n < x - u + \frac{\Delta}{2}\right) \quad (3)$$

We assume that measurements will be uniformly distributed over any given quantization interval. Shifting that interval to $(0, \Delta)$, we can find the expected distribution of pre-quantized values x after noise is added:

$$p(x)_{u+n} = p(x)_u * p(x)_n = -\frac{1}{2\Delta\sigma} \left(\operatorname{erf}\left(\frac{x-\Delta}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) \right) \quad (4)$$

The mass function of quantized values for a single pixel will be:

$$p(X) = \int_{-\infty}^{X+\frac{\Delta}{2}} p(x)_{u+n} dx - \int_{-\infty}^{X-\frac{\Delta}{2}} p(x)_{u+n} dx \quad (5)$$

The ROI mean distribution function will be the sum of an ensemble of many individual pixels distributed according to $p(X)$:

$$E[ROI] = \frac{1}{N} \sum_i P_i \quad (6)$$

where $P_i \sim p(X)$.

Although we have assumed pixel noise is independently distributed, the total intensity of each pixel in the ROI (signal + noise) will not be independently distributed. Given pixel intensities are not independently distributed, finding this sum analytically will be very difficult as the central limit theorem cannot be applied. Figure 2 shows the numerical solution of Equation (6) where total noise power as a function of white noise power constructively dithers the mean. Once the curve minimum is reached, total noise power increases linearly with random noise power. Figure 3 shows the effect of increasing gain, at each of the arbitrarily chosen five random noise levels. The reduction in total noise tends towards zero as gain increases until a critical point is reached where noise begins to increase again.

B. RANDOM NOISE

We expect to find random noise with power far outside of the range where increasing gain can effectively reduce total noise. Gain is applied after the exposure of the image sensor has finished, meaning that noise is amplified along with signal when increasing gain. We are then left with the parameter of exposure time and it's relation to the overall SNR. Increasing exposure time will increase the power of the signal relative to the noise as signal power is increasing over the entire duration of the exposure time. This provides an opportunity for SNR improvement after random noise level becomes higher than the maximum level for gain to improve total noise.

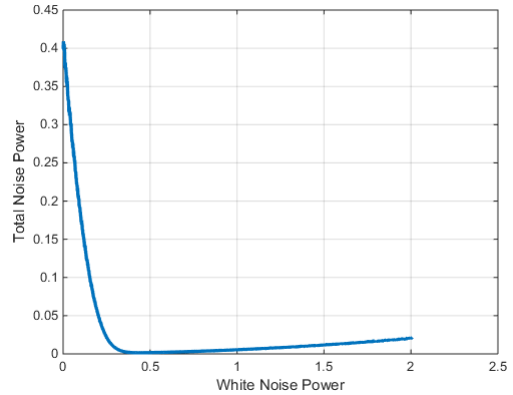


FIGURE 2. Total noise power (quantization + random) as a function of random noise power.

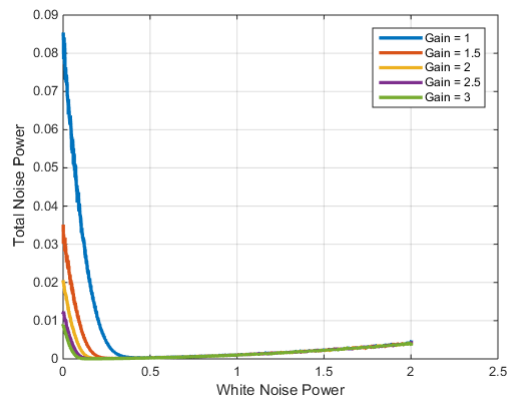


FIGURE 3. Total noise power over random noise power at different gain levels).

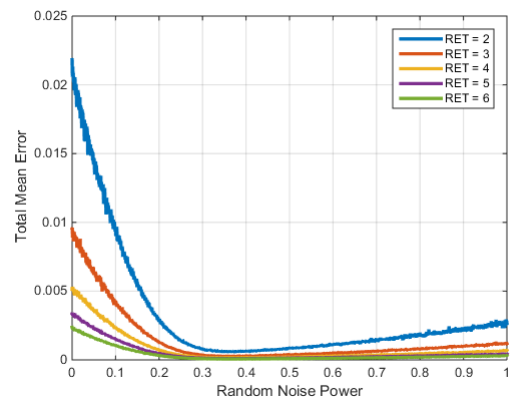


FIGURE 4. Total noise power over random noise power at increasing relative exposure times showing the continuing improvements provided by exposure time past the minimum.

Figure 4 shows that, unlike with gain, higher levels of exposure time continue to provide a net SNR improvement even as random noise increases past the noise minimum. Exposure time is expressed as a ratio relative to the initial exposure time, or Relative Exposure Time (RET). Figure 5 answers the question of combining gain and exposure time to achieve a total exposure (gain exposure time product). We see that as total noise increases past its minimum exposure time

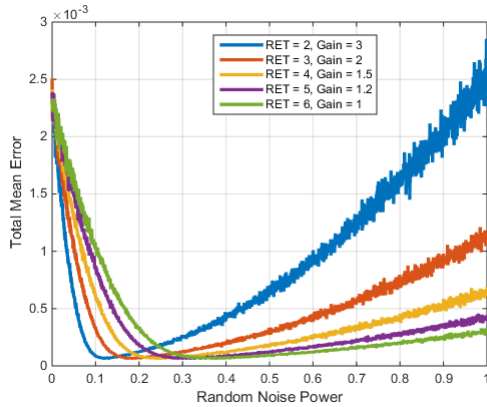


FIGURE 5. Total noise power over random noise power at the same total exposure (product of exposure time and gain) showing that maximizing exposure time is the most effective way to set exposure at realistic noise levels.

out-performs high gain significantly. As this higher noise level is closer to what we expect with real-world recordings, controlling exposure time will be the focus of our technique for overall exposure control.

C. EXPOSURE CONTROL

Our approach to extracting an rPPG is based on the distribution of pixel intensities in the ROI. If an infinite range of output pixel values were possible, then theoretically maximum SNR would be achieved from the maximum exposure time. However, in a real device, output values cannot continue to increase beyond the highest possible output and exposure is limited by pixel saturation. The ROI is also not expected to be perfectly homogeneous and therefore not all pixels will have the same intensity and some will saturate while others may not.

A predictable problem is distortion in the rPPG that is caused by pixel saturation in frames that sample close to the rPPG peak. To preempt this, we initially identify the frame that is closest to the peak of the rPPG by choosing the one with the highest ROI mean. A window of frames that is $\frac{4}{3}$ seconds long guarantees at least one rPPG peak. This is based on a desired HR estimation between 45-200 beats per minute (BPM) corresponding to frequencies between 0.75-3.33Hz. This then forms the basis for subsequent exposure time calculations.

Once a frame has been selected, mean and variance of the pixel values are then obtained from the ROI. As exposure increases, saturation of the higher valued pixels in the ROI will appear as soft-clipping distortion in the rPPG (Figure 6). The goal is to estimate exposure at the balancing point between preventing saturation of the pixels with highest intensities and causing an overall increase in the ROI pixel intensities. To achieve this we model the contribution of saturated and unsaturated pixels to the ROI mean.

The model assumes that pixel intensities in the ROI are normally distributed, this is equivalent to assuming the skin in the ROI is homogeneous with each pixel independently

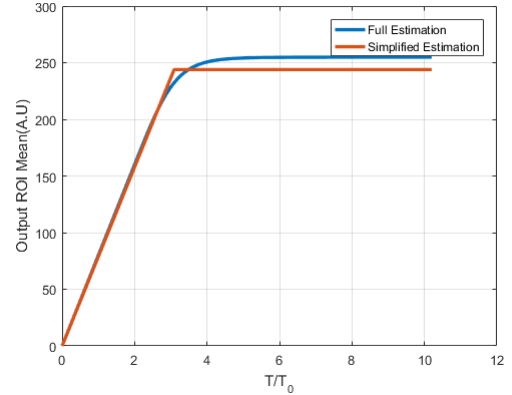


FIGURE 6. The full model ROI mean as exposure increases (blue) with the simplified model (orange).

affected by white noise. This implies that before any increase in exposure time is applied the distribution of pixel intensities in the ROI, $p(I)$, will be $- p(I) \sim \mathcal{N}(\mu, \sigma)$. The model further assumes that pixel intensity increases can be linearly approximated over small changes in exposure time, exposing the frame, including the ROI for some greater time T . This assumption will be correct over sufficiently small steps in exposure time, as it is mathematically correct if brightness level is differentiable. This will produce a new distribution $\frac{T}{T_0}p(I) \sim \mathcal{N}(\frac{T}{T_0}\mu, \frac{T}{T_0}\sigma)$ where T_0 is the current exposure time. Increasing exposure time has the effect of increasing variance of pixel intensities, and moving all pixels closer to the upper saturation point.

The ROI mean can then be modelled from the combination of unsaturated pixels and saturated pixels. Let M be the ROI mean, estimated as:

$$M = \mu'_T(1 - P_{sat}) + SP_{sat} \tag{7}$$

where S is the saturation value, μ'_T is the truncated mean:

$$\mu'_T = \frac{T}{T_0}\mu - \sigma \frac{\phi(q)}{\Phi(q)} \tag{8}$$

Here ϕ and Φ represent the normal probability density and cumulative density respectively, and P_{sat} is the probability of one given pixel in the ROI saturating:

$$P_{sat} = \Phi(z > q) \tag{9}$$

where $q = \frac{S - \frac{T}{T_0}\mu}{\frac{T}{T_0}\sigma}$ represents the probability of a pixel exceeding the saturation point. This is achieved by standardising the normal distribution after the exposure time has increased linearly by a factor of $\frac{T}{T_0}$. Truncation accounts for the removal of saturated pixels. The term SP_{sat} contributes to rPPG distortion, while the term $\frac{T}{T_0}\mu(1 - P_{sat})$ contributes to rPPG accuracy.

The model helps predict how exposure will produce distortion in the rPPG. The value of M as a function of $\frac{T}{T_0}\mu$ has a clear linear region and non-linear region (Figure 6). Overall non-linear amplification of a signal is a source of

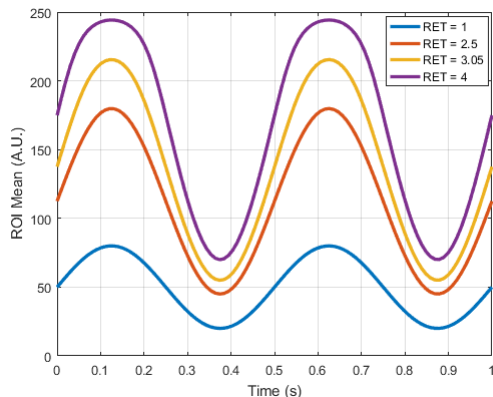


FIGURE 7. Sinusoidal ROI mean output over time at different exposure times.

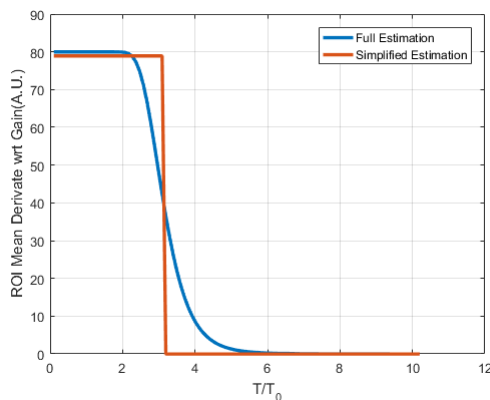


FIGURE 8. Derivative of the ROI mean with respect to exposure (both models).

distortion. rPPG distortion will occur when the product $\frac{T}{T_0} \mu$ for a frame reaches the point where M begins to enter the non-linear region. Figure 7 shows that as exposure time increases, the effective amplification will become non-linear. Maximum exposure time without distortion occurs when the rPPG peak reaches the end of the linear region. This can be shown once M is approximated as a piecewise function with a cutoff C (Figure 8). Let $\frac{T}{T_0} = A$:

$$M = \begin{cases} A\mu & A\mu \leq C \\ S & A\mu > C \end{cases} \quad (10)$$

Suppose there is some signal a with minimum a_0 and maximum a_1 then amplitude $A = a_1 - a_0$. If a is amplified according to the approximation of M the new amplitude, f_A will be:

$$f_A = \begin{cases} Aa_1 - Aa_0 & Aa_1 \leq C, Aa_0 \leq C \\ S - Aa_0 & Aa_1 > C, Aa_0 \leq C \\ 0 & Aa_1 > C, Aa_0 > C \end{cases} \quad (11)$$

S will differ from Aa_1 by some quantity δ . Substituting $S = Aa_1 + \delta$ and $Af = A(a_1 - a_0)$ gives:

$$f_A = \begin{cases} A & Aa_1 \leq C, Aa_0 \leq C \\ A + \delta & Aa_1 > C, Aa_0 \leq C \\ 0 & Aa_1 > C, Aa_0 > C \end{cases} \quad (12)$$

Therefore the condition for distortion free amplification is $Aa_1 \leq C$, or in general $A\mu_{MAX} \leq C$. This procedure for calculating maximum distortion-free exposure from an ROI can be summarised as follows:

- 1) Calculate P_{Sat} , μ'_T and σ'_T from the ROI
- 2) Estimate μ and σ from P_{Sat} , μ'_T , T and known current exposure time T_0
- 3) Calculate M over a range of possible exposures t
- 4) Find the cutoff of the linear region in M and its corresponding exposure time T_C
- 5) Set the current exposure time T_0 to T_C

1) CALCULATE P_{Sat} , μ'_A AND σ'_A FROM THE ROI

Pixel intensity statistics of *non-saturated* pixels within the ROI are calculated as follows:

$$\mu'_A = E[ROI_{p \neq S}] \quad (13)$$

$$\sigma'_A = Var[ROI_{p \neq S}] \quad (14)$$

$$P_{sat} = \frac{|ROI_{p=S}|}{|ROI|} \quad (15)$$

2) ESTIMATE μ AND σ FROM P_{Sat} , μ'_T , T AND KNOWN CURRENT GAIN T_0

To calculate the estimates of μ and σ from P_{Sat} , μ'_T and σ'_T the upper-tail truncation formulas are rearranged algebraically [24]:

$$\sigma = \frac{1}{T_0} \sqrt{\frac{\sigma'^2_T}{1 - q \frac{\phi(q)}{\Phi(q)} - (\frac{\phi(q)}{\Phi(q)})^2}} \quad (16)$$

$$\mu = \frac{1}{T_0} (\mu'_T + \sigma \frac{\phi(q)}{\Phi(q)}) \quad (17)$$

Values for $\phi(q)$ and $\Phi(q)$ are obtained from:

$$\phi(q) = \phi(\Phi^{-1}(P_{Sat})) \quad (18)$$

$$\Phi(q) = P_{Sat} \quad (19)$$

The above will only hold true if $P_{sat} < 0.5$, otherwise the truncation will no longer be exclusively in the upper tail and exposure time must be decreased until $P_{sat} < 0.5$ again.

3) CALCULATE M OVER A RANGE OF POSSIBLE EXPOSURES x

In our model, M contains sums and products of Gaussian and error functions. Due to the associated level of algebraic difficulty, it is better to numerically analyze M .

4) FIND THE CUTOFF OF THE LINEAR REGION IN M AND ITS CORRESPONDING EXPOSURE T_C

There are several methods available that may be used to select the end of a linear region. One simple approach is to base selection on a threshold in $\frac{\partial M}{\partial A}$. The derivative for the piecewise approximation of M is then:

$$\frac{\partial M}{\partial A} = \begin{cases} \mu & Aa_1 \leq C, A\mu \leq C \\ 0 & Aa_1 > C, A\mu \leq C \end{cases} \quad (20)$$

The cutoff is identified in the approximation as a discontinuity. Figure 6 demonstrates how $\frac{\partial M}{\partial A}$ will not have a discontinuity for M in equation (20). There will instead be a smooth transition from μ to 0 over the elbow of M as shown in Figure 8. Cutoff is then estimated to be at the level where $\frac{\partial M}{\partial A}$ crosses below a threshold relative to μ .

5) SET THE CURRENT EXPOSURE TIME T_0 TO T_C

The resulting exposure time T_C will deliver the best rPPG estimation. The final step is to set the current camera exposure time, T_0 , to T_C .

The initial assumption of linearity implies that an exposure time increase from T_0 to T will increase the ROI mean by a factor of $F = \frac{T}{T_0}$. In actuality, exposure for time T will produce an increase below F . Finding the true exposure time required to achieve an increase of F can be done with the algorithms discussed in the literature review. Taking this into account we then adjust the estimation of our desired exposure [23] time as:

$$T = \frac{(S - \mu)T_0T_M}{(F - \mu)T_0 + (S - F)T_M} \quad (21)$$

where:

$$T_M = \frac{S}{\mu}T_0 \quad (22)$$

This calculation must be made iteratively, but only on frames that contain peaks within a 1.5 second window as saturation distortion will occur in peak frames first. The process of calculating a desired ROI mean and then estimating the exposure time that will achieve it is incremented every 1.5 seconds until the desired result is achieved.

A final consideration is the potential effect that saturation could have on artifacts created by movement. Current rPPG methods assume that all artifacts created by movement are of equal size. However, if saturation is reached movement artifacts will become unequal between the channels. Uncorrected rPPG artifacts will then be significantly greater than the quantization noise we are trying to minimize. To mitigate this, as we calculate exposure time for all channels, then the lowest value is selected.

IV. IMPLEMENTATION

To test our approach for rPPG extraction we chose Feng's method [6]. On the first frame a face was detected with the VJ algorithm. The region spanning facial proportions (0.15, 0.35) to (0.25, 0.6) was chosen as the ROI. This area is in the centre of the forehead for convenience of homogeneity and size. Calculations to determine the rPPG were based on controlling the exposure of the ROI only. The face and ROI were tracked through subsequent frames using the KLT algorithm. For each colour channel in every frame an ROI mean of pixel intensities was calculated and formed three time series: $R(t)$, $G(t)$ and $B(t)$. These were subsequently combined using the technique described by Feng to create a

single estimate of PPG:

$$PPG(t) = \frac{R_f(t)}{I_r(t)} - \frac{G_f(t)}{I_g(t)} \quad (23)$$

where $R_f(t)$ and $G_f(t)$ are $R(t)$ and $G(t)$ that have been band pass filtered, while:

$$I_r(t) = \frac{R(t)}{\sqrt{R^2(t) + G^2(t) + B^2(t)}} \quad (24)$$

$$I_g(t) = \frac{G(t)}{\sqrt{R^2(t) + G^2(t) + B^2(t)}} \quad (25)$$

The bandpass filter applied to $R_f(t)$ and $G_f(t)$ passes only the band of reasonably expected Hearts Rates, often chosen as 0.75Hz to 4Hz (45BPM to 240BPM). This is the passband we used. A sliding 3 second window was taken from the exposed ROI to estimate HR from $PPG(t)$. At the centre of this window (1.5 seconds) frequencies as low as 0.75 Hz were able to be detected. Our approach to fundamental frequency estimation was to use autocorrelation [14] - the integral of the product of a signal with a delayed version of itself, $R_f(\lambda) = \int_{-\infty}^{\infty} f(x)f(x - \lambda)dx$. Although, all signals will have a high correlation at delay 0, periodic signals will have comparably high correlation at delays of all integer multiples of the fundamental frequency. Therefore, if signal S has period T it follows that $R_S(0) = R_S(T) = R_S(2T), R_S(3T) \dots$. The frequency of S can be estimated from consecutive peaks P_0, P_1 in it's autocorrelation with $f_F = \frac{1}{P_1 - P_0}$. More sophisticated approaches for processing the unstable instantaneous HR into a more stable output HR can be obtained by applying a moving average of window length six seconds but were not applied to the current algorithm [2], [25].

The iterative process of setting exposure time is performed at intervals of 45 frames. The calculations described above are performed to predict desired exposure time, the calculated exposure time is then compared with current exposure time. The cameras exposure time is iteratively increased until the calculated exposure time matches the current exposure time. This procedure will take much longer than the cameras in-built AEC, with each iteration requiring 1.5 seconds of data. However, the longer total time of execution for our novel algorithm is not significant compared the minutes or hours of superior performance it may provide.

V. EXPERIMENTAL VALIDATION

We designed two experiments to answer the following questions: (i) How does the novel algorithm change the performance of HR estimation compared to a set of exposure times sampling the available range?; (ii) How does the novel algorithm change the performance of HR estimation compared to automatic exposure time that is inbuilt on a research camera?

Participants in both experiments were seated 3 metres from a two camera configuration. This testing distance will mean the ROI will occupy a very small area of the frame, less than 1%. The small relative size of the ROI is key to needing the camera to be placed close to the subject. This is similar to the distance that would be seen in real world deployments.

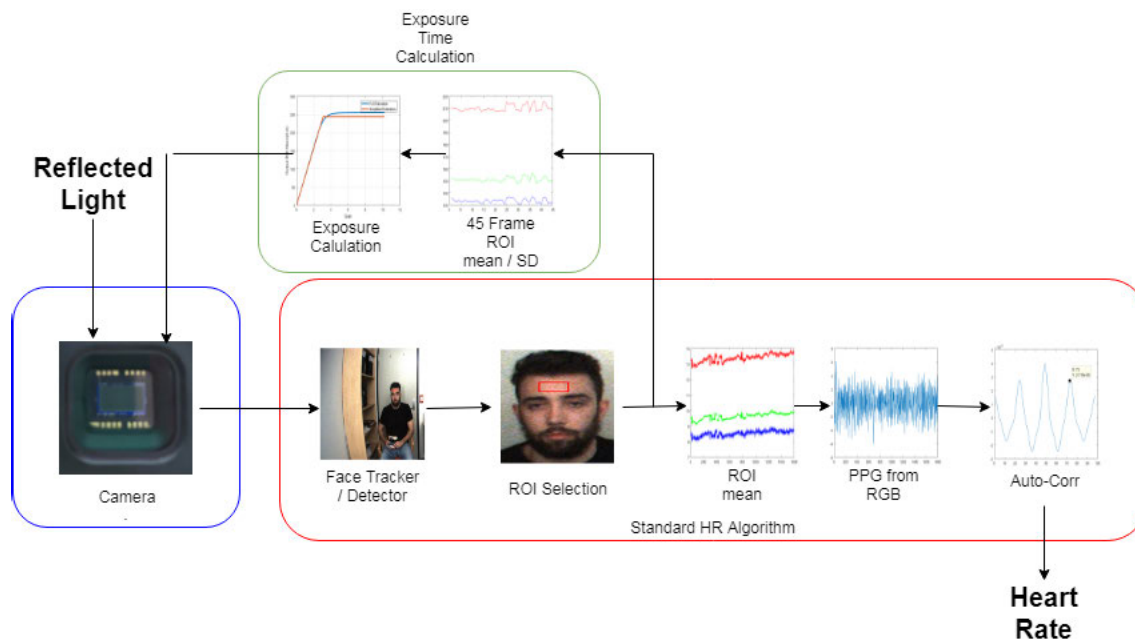


FIGURE 9. An overview of the full HR estimation system. Periodic increments are made to exposure time to the image sensor of Camera A based on ROI statistics until current exposure time is closest to the predicted best exposure time. Exposure time of the image sensor in Camera B is not adjusted.

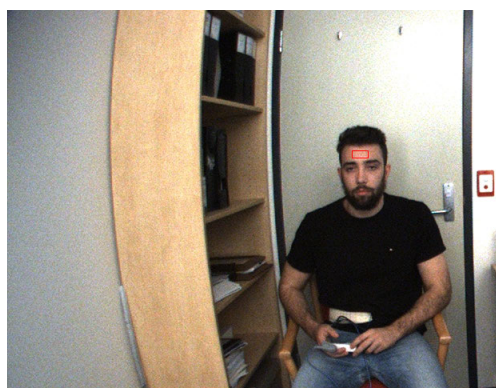


FIGURE 10. Experimental set-up in an office with standard fluorescent lighting, with the ROI for rPPG shown.

A research camera with inbuilt AEC bases its calculations on the whole frame and implies information obtained from the ROI will be less effective at a distance. Two identical *Point Grey Flea3* cameras were used to obtain the mean absolute error, standard deviation of absolute error and time spent within 6 BPM relative to ground truth provided from a Compumedics[®] Sompté clinical monitor routinely used for ambulatory sleep studies to record contact PPG and HR. Gain for both cameras was set to 1 (0 dB) as this gave maximum room for exposure increase without distortion. Ethical approval was granted by the QUT Human Research Ethics Committee (HREC)

A. EXPERIMENT 1

Consent was obtained from five participants enrolled in this experiment. Exposure time on Camera A was iteratively

increased by 1ms. Both cameras recorded with a frame rate of 30fps. A factory set resolution of 480 × 620 was also chosen and although not a standard resolution it had no effect on calculations because of the size and location of the ROI in the camera frame. One minute of synchronised data was gathered on both cameras for each of the predetermined exposure times tested on Camera B. This was repeated for each of the five participants.

B. EXPERIMENT 2

Consent was obtained from ten participants enrolled in this experiment. This experiment was smaller, designed specifically to test our novel algorithm against the camera’s in-built AEC only. Exposure time on Camera A was iteratively increased by 2ms, Camera B had exposure time set by the cameras in-built AEC. Both cameras recorded with a frame rate of 30fps. Again, a resolution of 480 × 620 was also chosen. One minute of synchronised data was gathered on both cameras for each of participant.

VI. RESULTS

A. EXPERIMENT 1

Results from the static exposure time and our algorithmic exposure time comparison recordings are shown below in Table 1. Trends in MAE and 6BPM time over all participants are shown in Figures 11 and 12. The x-axis in these figures corresponds to static exposure time in each recording.

Table 1 shows the per recording performance of HR estimation with our novel algorithm against a set of static exposure times. The cameras inbuilt AEC was briefly tested before recording for each participant, it selected an exposure of

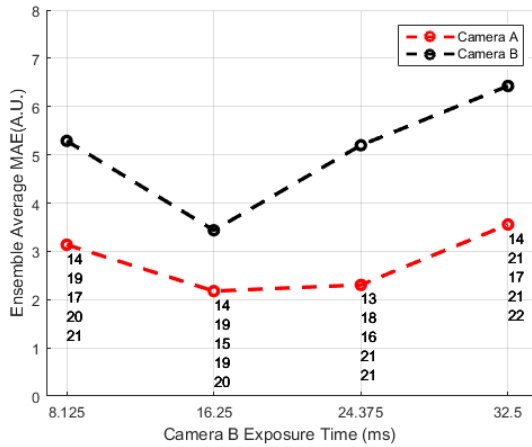


FIGURE 11. Comparison of camera performance based upon mean absolute error from the chosen ground truth. Camera A data point labels indicate the corresponding exposure times for each of the 5 participants in milliseconds for each of the arbitrarily chosen exposure times for Camera B.

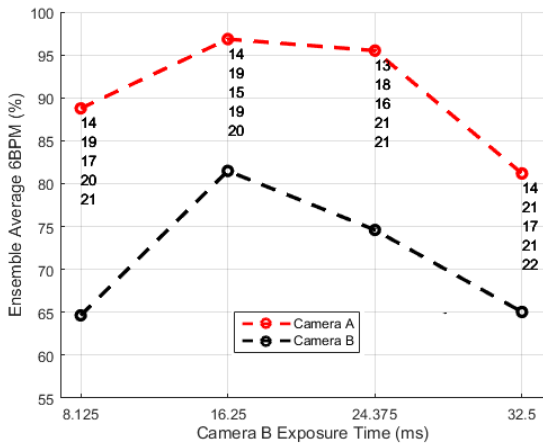


FIGURE 12. Comparison of camera performance based upon percentage of estimated HR within 6 BPM of the chosen ground truth. Camera A data point labels indicate the corresponding exposure times for each of the 5 participants in milliseconds for each of the arbitrarily chosen exposure times for Camera B.

12ms for all 5 participants. Superior performance for each recording is highlighted in green, with inferior performance highlighted in red. Results demonstrate that Camera A with the novel algorithm out performed Camera B for 4 out of 5 participants. Camera A demonstrated a mean improvement of 1.62 BPM, 2.62 BPM and 3.46 BPM over Camera B at each of the pre-set conditions of over exposed, under exposed and well exposed recordings respectively.

B. EXPERIMENT 2

The results presented in Table 2 demonstrate that the novel algorithm set the exposure time to be longer for all of the participants studied. On average the novel algorithms chosen exposure time was 11.32ms longer. The results also show that Camera B with the inbuilt automatic exposure varied less across participants, with standard deviation of 3.48ms

TABLE 1. Comparison of heart rate estimation error and percentage agreement within 6 BPM of ground truth at each Camera B exposure time. Superior and inferior performance are highlighted in green and red respectively. MAE = Mean Absolute Error, 6BPM Time = Proportion of time withing 6BPM, ET = Exposure Time.

Camera A (Novel Algorithm)			Camera B (Pre Set)		
ET (ms)	MAE (BPM)	6 BPM (%)	ET (ms)	MAE (BPM)	6 BPM (%)
Subject 1					
14	2.17	98.6	8.125	6.94	52.8
14	2.84	99.5	16.25	4.65	66.9
13	3.85	81.9	24.325	12.05	44.1
14	2.82	97.7	32.5	10.7	51.3
Subject 2					
19	2.57	92.35	8.125	3.49	80.8
19	2.72	89.9	16.25	3.6	82.8
18	2.35	98.2	24.325	3.98	75
21	3.59	77.9	32.5	5.95	66.9
Subject 3					
17	2.01	100	8.125	4.68	65.3
15	2	196.2	16.25	3.85	83.3
16	1.64	100	24.375	4.1	77.3
17	4.29	70.7	32.5	6.83	58.1
Subject 4					
20	5.93	62.9	8.125	8.33	44.3
19	1.46	98.6	16.25	2.5	87.33
21	1.75	100	24.325	3.68	76.9
21	4	76	32.5	4.37	79.3
Subject 5					
21	5.34	56.8	8.125	4.7	63.8
20	1.86	100	16.25	2.58	87.1
21	1.93	74	24.325	2.18	99.7
22	3.08	83.6	32.5	4.28	69.6

TABLE 2. Comparison of heart rate estimation error and percentage agreement within 6 BPM of ground truth using Camera B factory algorithm for automatic exposure. Superior and inferior performance are highlighted in green and red respectively. MAE = Mean Absolute Error, 6BPM Time = Proportion of time withing 6BPM, ET = Exposure Time.

Subject	Camera A (Novel Algorithm)			Camera B (Automatic)		
	ET (ms)	MAE (BPM)	6BPM Time (%)	ET (ms)	MAE (BPM)	6BPM Time (%)
1	23.6	2.08	100	14.7	2.66	98.1
2	27.9	1.67	100	13.5	4.38	75.8
3	23.8	5.25	60.3	11.7	10.97	48.3
4	21.9	5.2	66.8	14.3	5.77	53.8
5	22.9	2.48	100	11.6	4.87	72.2
6	22.1	2.94	84.8	11.2	10.96	27.6
7	27.6	5.02	70.5	12	5.38	64.8
8	20	4.59	73.2	5	12.41	62.1
9	14.1	4.18	79.8	6.6	7.81	35.4
10	16.2	5.75	54.7	6.3	10.2	37.6

compared with the more variable novel algorithm with standard deviation 4.38ms. The novel algorithm performed better across all participants with a mean improvement in 3.625 BPM mean absolute error and an overall improvement of 21.44% in heart rate estimations that were within 6 BPM of the chosen ground truth. The decrease in mean MAE is statistically significant under t-test with null hypothesis H_0 : our novel algorithm's mean MAE is not less than the AEC mean MAE, with $p = 0.0045$. The increase in mean 6BPM time is statistically significant under t-test with null

hypothesis H_0 : our novel algorithms mean 6BPM time is not greater than the AEC mean 6BPM time, $p = 0.0118$.

C. EXECUTION TIME

The novel algorithm required an average of 15.15 seconds of execution time in the first experiment and 21.015 seconds of execution time in experiment two. The execution time of our novel algorithm was very slow as compared to the cameras factory AEC, which completes on the order of milliseconds. The vast majority of the execution time required by our algorithm was for gathering data, the time of actual computation between the two algorithms is very similar.

VII. DISCUSSION

The results from both experiments clearly demonstrated that our novel algorithm outperformed a set of sample exposure times across the range of exposure times available on a research camera including the factory algorithm for automatic exposure when used for the purpose of estimating heart rate. In the first experiment our novel algorithm produced superior performance in HR estimation over a range of exposure times and subjects. In the second experiment our novel algorithm outperformed the cameras in-built exposure algorithm on all ten subjects. These two experiments provide strong evidence that our novel exposure time algorithm chooses exposure times that are both overall effective and superior than a cameras default algorithm for the purpose of heart rate estimation from rPPG.

It was observed that performance in the final recording of each participant in experiment 1 was lower even with the algorithmic exposure. This may be because of increased movement from the subjects becoming more restless after they have been sitting for about 10 minutes by the end of the fourth recordings. It is also notable that the default AEC chose a much more consistent exposure time across all participants, with only 3.48ms of standard deviation compared to our algorithms 4.38ms: this is explained by the cameras default AEC basing calculations off the entire frame.

An interesting observation was the variation of algorithm selected exposure time for different subjects. Camera B inbuilt AEC in experiment 1 selected the same exposure time across all participants at 12ms. In the second experiment, where Camera B used the factory algorithm for automatic exposure, more variation was observed in that camera's determined exposure time compared to the constant 12ms observed in experiment 1. This was due to the different environmental conditions where natural light was introduced with the fluorescent lighting as experiment 2 was conducted in a room with a large window. The purpose specific algorithm determines its exposure time based only upon the chosen ROI. This demonstrates how purpose specific exposure control could be applied to a range of circumstances: when the subject is far away, when the scene is unevenly lit or when skin colour is highly contrasted from the background but occupies only a small proportion of the frame.

For any given subject, there was minor variability of output exposure time from our algorithm across different recordings for each participant, with a variance of around 1.5ms. Considering the combination of very slight environmental changes between recordings and iteration interval being 1ms, this inter-recording variation may be due to the true exposure time for highest SNR may lie near the middle of an interval, where a slight change in value will change the iteration output by 1.

Limitations of our novel algorithm currently are execution time and subject movement. It was stated earlier that the algorithm had the ability to estimate the highest SNR, distortion free exposure time, despite the fact that it would be iteratively set in our specific experiment. Operating as is, our novel algorithm is much slower than the cameras inbuilt AEC. This cannot be avoided as setting exposure time specifically requires collecting a full period of HR data, which will always take on the order of seconds. However, we believe an adjustment period of seconds is acceptable when compared to the performance increase our algorithm provides. These improvements would continue over the duration of usage, which in a clinical setting could extend over minutes or hours. Although subject movement was not an issue in our experiment it is unclear how excessive movement may effect our novel algorithm if it was present. Our algorithm will place the upper portion of the colour channels closer to saturation than they would otherwise be, this leads to superior performance without excessive movement present, but may actually lead to worse performance when there is a large amount of subject movement.

VIII. CONCLUSION

This paper attempted to address the lack of investigation of techniques for potential improvement of existing methods of HR extraction from standard video. We proposed a method for automatically controlling exposure in RGB cameras specifically for the purpose of rPPG. We proposed a method to control exposure time specifically for rPPG by estimating the highest exposure time possible without saturation occurring in the ROI. Our experiments tested the difference in error between two standard research cameras relative to a standard pulse oximeter, one camera with a static exposure times and the other operating our novel algorithm. This experiment was conducted across five people, for each the algorithm was compared to four different static exposure times sampling the range of possible exposure times. Our algorithm produced a lower error across in the best case for 4 out of 5 participants. The agreement between the exposure time that was arrived at iteratively and predictions by the algorithm also provides evidence that the iterative component could be replaced by prediction in future work. Future work may also include experiments with lower frame-rates as a method to provide the opportunity for much greater exposure times for very dark environments. Finally, future work may also focus on testing this system in challenging out of lab environments.

REFERENCES

- [1] A. Al-Naji and J. Chahl, "Remote optical cardiopulmonary signal extraction with noise artifact removal, multiple subject detection & long-distance," *IEEE Access*, vol. 6, pp. 11573–11596, 2018.
- [2] F. Andreotti, A. Trumpp, H. Malberg, and S. Zaunseder, "Improved heart rate detection for camera-based photoplethysmography by means of Kalman filtering," in *Proc. IEEE 35th Int. Conf. Electron. Nanotechnol. (ELNANO)*, Apr. 2015, pp. 428–433.
- [3] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Amplitude-selective filtering for remote-PPG," *Biomed. Opt. Express*, vol. 8, no. 3, pp. 1965–1980, 2017.
- [4] G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiological Meas.*, vol. 35, no. 9, pp. 1913–1926, Sep. 2014.
- [5] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
- [6] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.
- [7] B. D. Holton, K. Mannapperuma, P. J. Lesniewski, and J. C. Thomas, "Signal recovery in imaging photoplethysmography," *Physiological Meas.*, vol. 34, no. 11, pp. 1499–1511, Nov. 2013.
- [8] R.-Y. Huang and L.-R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," *Biomed. Eng. OnLine*, vol. 15, no. 1, p. 11, Jan. 2016.
- [9] W.-C. Kao, L.-W. Cheng, C.-Y. Chien, and W.-K. Lin, "Robust brightness measurement and exposure control in real-time video recording," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 4, pp. 1206–1216, Apr. 2011.
- [10] F.-T.-Z. Khanam, A. Al-Naji, and J. Chahl, "Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: A review," *Appl. Sci.*, vol. 9, no. 20, pp. 4474–4506, 2019.
- [11] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Express*, vol. 6, no. 5, p. 1565, 2015.
- [12] J.-S. Lee, Y.-Y. Jung, B.-S. Kim, and S.-J. Ko, "An advanced video camera system with robust AF, AE, and AWB control," *IEEE Trans. Consum. Electron.*, vol. 47, no. 3, pp. 694–699, Aug. 2001.
- [13] G. Madhavan, "Plethysmography," *Biomed. Instrum. Technol.*, vol. 39, pp. 367–371, 2005.
- [14] P. Mcleod and G. Wyvill, "A smarter way to find pitch," in *Proc. Int. Comput. Music Conf.*, 2005, pp. 138–141.
- [15] B. D. Nam, M. H. Cho, and S. G. Lee, "The fast auto exposure algorithm based on the numerical analysis," *Proc. SPIE*, vol. 3, 93–99, 1999.
- [16] N. Nourani-Vatani and J. M. Roberts, "Automatic camera exposure control," in *Proc. Australas. Conf. Robot. Autom.*, 2007, pp. 1–6.
- [17] S. Nuske, J. Roberts, and G. Wyeth, "Extending the dynamic range of robotic vision," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2006, pp. 162–167.
- [18] S. Nuske, J. Roberts, and G. Wyeth, "Visual localisation in outdoor industrial building environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2008, pp. 544–550.
- [19] R. Sahni, "Noninvasive monitoring by photoplethysmography," *Clinics Perinatology*, vol. 39, no. 3, pp. 573–583, Sep. 2012.
- [20] S. Schulz and M. Grimm, "Using brightness histogram to perform optimum auto exposure camera," *Tech. Rep.*, 2007, vol. 2, no. 2.
- [21] I. Shim, T.-H. Oh, J.-Y. Lee, J. Choi, D.-G. Choi, and I. S. Kweon, "Gradient-based camera exposure control for outdoor mobile platforms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1569–1583, Jun. 2019.
- [22] J. Smith and R. Roberts, *Vital Signs for Nurses: An Introduction to Clinical Observations*. 2011, pp. 61–81.
- [23] Y. Su, J. Y. Lin, and C.-C. J. Kuo, "A model-based approach to camera's auto exposure control," *J. Vis. Commun. Image Represent.*, vol. 36, pp. 122–129, Apr. 2016.
- [24] C. Tomasi and T. Kanade, "Shape and motion from image streams: A factorization method-part 3 detection and tracking of point features introduction," *Tech. Rep.*, Apr. 1991.
- [25] D. Nhan Tran, H. Lee, and C. Kim, "A robust real time system for remote heart rate measurement via camera," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jun. 2015, pp. 1–6.
- [26] 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.
- [27] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, p. 1.
- [28] W. Wang, S. Stuijk, and G. de Haan, "Unsupervised subject detection via remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2629–2637, Nov. 2015.



JORDAN LAURIE received the B.Eng. degree in electrical from the Queensland University of Technology, where he is currently pursuing the Ph.D. degree. His research interest includes vital signs assessment using cameras.



NIALL HIGGINS is currently a Conjoint Senior Research Fellow with the Royal Brisbane and Women's Hospital and the Queensland University of Technology. His Ph.D. was based on telecommunication techniques using video in a clinical setting. He has over 20 years of clinical experience in acute nursing care including emergency and in mental health.



applications to industries such as mining, aerospace, or medical.

THIERRY PEYNOT (Member, IEEE) received the Ph.D. degree from LAAS-CNRS, University of Toulouse, France. He is currently an Associate Professor and the Mining3 Chair in robotics with the Queensland University of Technology (QUT). He is a Former Research Fellow with the Australian Centre for Field Robotics, University of Sydney. His main research interests include field robotics, multisensor resilient perception in challenging environments, and robotic vision, with



JOURNAL OF ROBOTICS AND AUTOMATION LETTERS.

JONATHAN ROBERTS (Senior Member, IEEE) was the Research Director of the CSIROs Autonomous Systems Laboratory. He is currently a Professor in robotics with the Queensland University of Technology (QUT). He is a Past President of the Australian Robotics Automation Association Inc. His main research interests include field robotics, medical robotics, and aerial robotics. He is a member of the Founding Editorial Board. He is also a Senior Editor of the IEEE

• • •