Direct-Global Separation for Improved Imaging Photoplethysmography

Jaehee Park , Ashutosh Sabharwal , Ashok Veeraraghavan
Rice University
6100 Main St, Houston, TX 77005
[jaehee.park, ashu, vashok]@rice.edu

Abstract

Camera-based estimation of vital signs has made significant progress in last few years. Despite of the significant algorithmic advances, the low signal-to-background ratio in video-based photoplethysmography continues to be a performance bottleneck. One of the main challenges is that much of the light returning to the camera from the subject is surface reflection from the skin and other dermal layers, and hence does not contain any pulsatile blood perfusion information to estimate photoplesthysmogram (PPG). In this paper, we show that direct-global separation techniques designed to reject much of the surface reflection photons can improve the signal-to-background ratio in the raw captured video signal. We study two techniques for the suppression of direct surface reflection (a) cross-polarization and (b) structured illumination. Using a dataset from 28 participants, our results show an average SNR improvement in estimating PPG from the use of structured illumination is 1.42 dB compared to the brightfield illumination. The use of cross-polarizers leads to an average SNR increase of 1.49 dB compared to brightfield illumination. And the combined structured illumination and polarizer method increases the SNR on the average by 1.90 dB compared to the brightfield illumination. The key result is that local PPG estimate SNR can increase to more than 5.63dB, enabling very large gains on regions with a large specular component. The RMSE decreased 55% *and the range of error reduced by* 12.9% *with the use of a polarizer and structured illumination.*

1. Introduction

Vital signs offer a way to continuously monitor the health or alertness of a person. In many controlled scenarios such as hospitals, contact devices such as pulse oximeters and electrocardiographs (ECG) is the de-facto standard of care. Such contact-based devices provide accurate, robust measurements.

However, there are several other application domains wherein, it is desirable to measure vital signs in a non-contact manner because contact with the skin should be avoided or is infeasible. Examples include vital signs monitoring on sensitive populations such as neonates and burn victims, in which contact increases the possibility of infection. Other scenarios include non-medical applications such as entertainment or driver monitoring—applications in which contact-based measurements are infeasible due to practical constraints.

Over the last few years, significant improvements have been demonstrated in algorithms for camera-based remote photoplethysmography (rPPG) estimation [1, 2, 3, 4, 5, 6, 7, 8]. These advances have enabled non-contact measurement of vital signs, such as heart rate (HR) and heart rate variability (HRV), with accuracy comparable to that of contact-based devices.

In spite of these significant advances, one key challenge remains: much of the light that is recorded at the sensor comes from surface reflection and sub-surface reflection components that did not interact with blood. As a consequence, the pulsatile signal that is critical for estimating PPG is buried in a large background signal which is composed primarily of surface reflections. This background signal comes from light that does not penetrate deep enough into the skin and thus does not interact with the blood vessels. Furthermore, the part of the signal that contains pulsatile information is extremely low and this problem is exasperated in low lighting conditions (like in NICUs) and in darker skin tones.

The key idea in the paper is to exploit two computational illumination and imaging techniques: (a) cross-polarization and (b) structured illumination, to reject much of the surface reflection photons increasing the signal quality for PPG estimation. We are the first to analyze small time-signal estimation from a video with cross-polarization and structured illumination based global separation; here the small signal is the PPG waveform that is buried in the large surface reflection component.

The overall system consists of a "front-end" illumination control combined with an algorithm to extract iPPG from the video of the exposed skin surface. The front-

end consists of either structured illumination or combined structured illumination and polarizer to first separate direct-global components. We have chosen direct-global separation methods that can operate on a per-frame basis, as our objective is to extract a small-magnitude time-domain PPG signal from the videos.

We collected data from 28 adult subjects with ground truth PPG using a contact-based pulse oximeter. The data was collected with structured illumination, polarizer combined with structured illumination and brightfield illumination. Our key results are as follows.

1. The average SNR increase in the estimation of iPPG, using global-separation techniques compared to the standard brightfield illumination is 1.60 dB, with range of improvement from 1.25 dB to 2.00 dB.

2. The average SNR improvement was almost constant as a function of skin tone, specifically the improvement numbers were 1.62 dB for light, 1.70 dB for medium, and 1.50 dB for dark skin tones.

3. While the average improvement is computed over the whole exposed region, the improvement in the SNR of specific skin regions (with high specular component) ranged from 3.48 dB to more than 5.63 dB. This large local improvement can have a significant impact on methods that aim to measure local perfusion, see e.g.[9].

4. The increase in SNR translated into a decrease in RMSE for heart rate estimation by using global separation. The RMSE decreased 55% from 0.3635 to 0.1680 with the use of polarizer and structured illumination. The error range reduced by 12.9%.

The rest of the paper is organized as follows. In section 2, we discuss prior works on remote vital signs estimation as well as two direct-global separation methods – cross-polarization and structured illumination. In section 3, we explain how direct-global separation improves the iPPG signal quality with a signal-to-background ratio. Section 4 explains the pipeline of acquiring videos using standard brightfield illumination, structured illumination, and cross-polarizers. We also define the algorithm used to extract the iPPG signal. In section 5, we describe the data acquisition protocol and experimental setup. Then in section 6, we present our result.

2. Prior Methods

We provide a brief overview of prior methods both for remote estimation of vital signs and for direct-global separation.

2.1. Remote Vital Signs Estimation

Over the last decade, researchers have been trying to estimate vital signs remotely using a non-contact based system such as a camera due to its low cost and ease of use[10][11]. The main drawback with using a camera is that the strength of pulsatile PPG signal is very low mainly because the blood volume change is small in intensity [12]. The signal containing the pulsatile information is small in contrast to the background. In addition to the challenge of detecting the low-intensity signal, the image capturing process contaminates the signal with photon shot noise, camera quantization noise, and read noise. The skin tone of the subject also affects the signal strength which decreases with higher melanin content in the skin that absorbs a large amount of incident light. Small movements of the subject like talking and nodding can cause large changes in the skin surface reflectance captured by the camera, affecting the quality of the estimated signal.

To reduce the effects of noise and motion, algorithms have been used to filter out the unwanted artifacts [4, 13, 14, 15]. These algorithms can be classified broadly into two categories–signal processing based and machine learning-based approaches. It has been previously studied in [3],[16] that the pulse rate (PR) and the breathing rate (BR) can be estimated with the help of a color camera and ambient illumination. Simple band-pass filtering and Fast Fourier Transform (FFT) was used to estimate the PR and BR from the extracted PPG signal. Independent Component Analysis (ICA) have been used to decompose the red, green, and blue channels into three independent sources [1],[17]. The PPG signal is then extracted using the separated green channel which can yield better PPG estimates.

In a later study, Principal Component Analysis (PCA) was used to extract PPG signal from a standard webcam. To eliminate the effects of motion artifacts, the authors in [2] have used a chrominance based method to reliably extract heart rate using a camera from exercising subjects, and have shown that this method works better than ICA and PCA for both stationary and moving people. Further, the adaptive LMS filter was used to reduce the motion artifacts from a corrupted PPG signal [18]. They have used FFT, SVD, and ICA to generate a noise reference signal, then applied an adaptive step-size least mean squares (ASLMS) filter for estimating an artifact-reduced PPG signal. Recently, a supervised learning based approach was proposed in [19] to estimate the heart rate after extracting PPG signal obtained from an off-the-shelf webcam. A trained SVM model is applied as a sliding window on the extracted noisy PPG signal to filter out false beats introduced by noise. Authors in [20] have used Support Vector Regression (SVR) on chrominance based method obtained from a video camera.

Almost all of the discussed methods is a post-processing

based approach aimed at reducing the effects of artifacts in estimating vital signs from the PPG signal. Rejecting noise in the data-capturing and pre-processing step could increase the performance of the algorithm as well as the post-processing evaluation. The strength of the light containing the pulsatile information that enters the camera is small compared to the background signal. Eliminating the direct surface reflection that does not contain pulsatile information can enhance the PPG signal strength as discussed in the next section.

2.2. Direct-Global Separation

When a face or other exposed skin area of the human body is lit by a source of light, and a camera image acquired, the intensity recorded at each pixel can be viewed as having two components, namely, direct and global. The direct component is due to the direct illumination of the point by the source and is typically caused due to surface reflection. The indirect (or global) component in such a scenario is predominantly composed of the sub-surface reflection components, i.e., the photons that penetrate the skin and interact with deep tissue and then exit the skin after multiple scattering events. In the context of remote PPG estimation, it is these sub-surface scattered global photons that contain information about the pulsatile component. Unfortunately, the direct surface reflection photons are much more numerous than the sub-surface scattered photons and this typically limits the signal to background ratio in camera-based vital signs estimation.

Over the last two decades, several computational illumination and imaging-based techniques have been developed for separating the direct and the global components of the image intensity. Two of the most popular and effective techniques are (a) cross-polarization and (b) structured illumination; both explained next.

2.3. Cross-Polarization

When light interacts with a surface like the human skin, the surface reflection components retain the polarization state of incident light. In contrast, the sub-surface scattering components have, through multiple scattering interactions, completely lost their polarization state and are randomly polarized. This difference can be used to enhance the sub-surface signal component. As an illustrative example, assume that the illumination source is horizontally polarized. The surface reflection retains the polarization state and so remains predominantly horizontally polarized. In contrast, the sub-surface component has lost its polarization state and so contains equal parts horizontal and vertical polarized light. If we add a vertical polarizer in front of the camera sensor, then most of the horizontal polarized surface reflection is rejected – thereby significantly improving the sub-surface or global component.

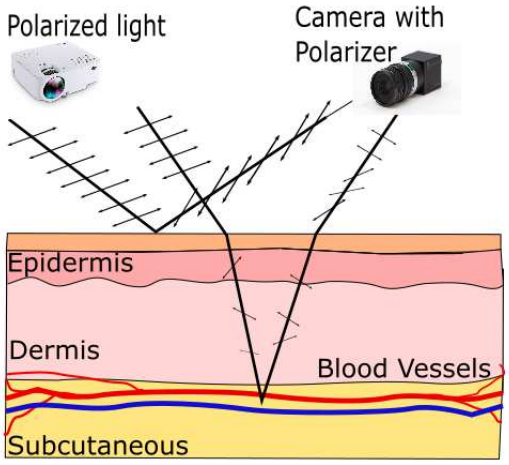


Figure 1: An illustration of on how the cross polarization works. When polarized light hits the surface of the skin without penetrating the skin, polarization is retained. When polarized light penetrates the skin, the light loses its polarization.

There are many applications in using cross-polarizers from synthesizing novel images to capturing hidden structures in medical imaging. Specular reflections are polarized while diffuse reflections are unpolarized [21]. Additionally, the diffuse amount of polarization depends on the angles of incidence and reflection [21]. Using these known properties of polarized light, images were captured with cross-polarized cameras from multiple directions in a controlled lighting environment to understand the skin reflectance property [22]. With the image collection of skin appearance with varying polarization, the authors were able to reconstruct realistic looking subsurface scattering representation of skin [22].

Characterizing skin appearance with cross-polarizers has been studied in biomedical systems as well. Polarizers have been used with medical imaging systems such as confocal fluorescence microscopy as well as conventional RGB cameras for acquiring PPG signals [23]. The depolarization of backscattered light that penetrates into the tissue depends on the scattering characteristics of the tissue [24]. Studying how polarized light interacts with different types of tissues can help non-invasively characterize tumors [25]. The point spread function of reflected polarized light has been studied to aid in the application of using cross-polarizers for imaging superficial tissues [26]. Clinical studies demonstrate that a camera with cross-polarizers could identify a more accurate representation the size and margins of lesions and basal cell carcinomas than a doctor could with her unaided eye [25]. Recent work by Sidorov et. al and Trumpp et. al show experimentally that polarizers benefit camera-

based photoplethysmography applications [27],[28]. Although polarizers are effective in reducing specular reflections, it also decreases the intensity of light that passes through the polarizer by half. Reducing the number of photons that reach the camera is a limitation especially in low light scenarios. Structured illumination is an alternative to separate the global components of light without losing the light that reaches the camera.

2.4. Structured Illumination

Consider a simple example wherein a structured illumination pattern is projected on a surface, such as human skin, the resulting image acquired on an image sensor. The spatial pattern projected on the skin can be decomposed into its frequency components. The surface reflection (or direct reflection) acts as an all-pass filter letting all spatial frequencies to be reflected back towards the camera. In contrast, the sub-surface scattering acts as a low pass filter, and only the low spatial frequencies are retained – the higher spatial frequencies are attenuated by the multiple scattering events. This difference in how the low and high spatial frequencies are attenuated by the direct surface reflection component and the indirect sub-surface component can be utilized to separate the direct and the global components.

Nayar et. al [29] demonstrated that capturing two images with a dense binary illumination pattern L and its complement $\bar{L}$ can separate the scene into global and direct components of light. The irradiances L^i and $\bar{L^i}$ are compared for a scene point S^i. If S^i is lit under a high-frequency pattern, the irradiances can be written as:

$$\begin{aligned} L^i &= L_d^i + \gamma L_g^i \\ \bar{L^i} &= (1-\gamma) L_g^i \end{aligned} \tag{1}$$

where L_d^i and L_g^i are the direct and global components of the irradiance at a scene point S^i and the γ is the fraction of activated source pixels. If we know which patch is lit directly by the source in the first image and which is not lit in the second image, we know the γ term and can compute the direct and global components at each camera pixel with the two images. Nayar et. al demonstrated several patterns, one of which was using checkerboard illumination shifts in which a number of images are taken with shifted checker patterns. The maximum and minimum measured brightness found at each pixel are used to compute the direct and global estimates. High-frequency sinusoidal patterns that vary over space and/or time can be used for direct-global separation using three patterns. The brightness of the projector pixel for the first pattern is generated by using a uniform distribution between 0 and 1. Two more patterns are generated by changing the phases of the sinusoids.

This method was later improved by allowing structured light algorithms to work in dynamic scenes [30]. Direct-global separation has also improved the performance of structured light 3D reconstruction that excludes global effects [31]. In medical applications, mapping of tissue optical properties with modulated imaging has been studied to characterize diffusive systems [32].

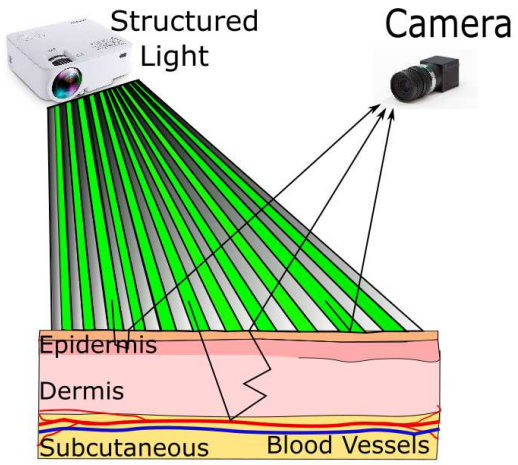


Figure 2: An illustration of the workings of structured illumination. High-frequency binary pattern is projected onto the skin. The areas of the skin directly lighted contain only the surface reflections whereas the areas that are not directly lighted contain both the global and direct component of light. Direct reflection acts as an all-pass filter whereas sub-surface scattering acts as a low pass filter.

Single image global separation has been explored for applications where the scene is not static and would benefit from using all frames in a video. A single-image method using high-frequency striped patterns to separate global components was presented by Nayar et. al [29].To separate the global component using a single image, first, assign a pixel a maximum or minimum label in a window around it. Then, the pixel intensity values at the maximum and minimum values are interpolated to obtain the L_{max} and L_{min} at the full resolution where L_{max} is the L^i and the L_{min} the complement $\bar{L^i}$. Recently, a real-time direct-global separation method has been developed that uses the relation between stereo geometry and light transport [33]. The integration of global-separation into medical applications such as ours for PPG estimation could be realized as more robust and efficient techniques are developed for global light imaging.

3. Direct Global Separation for Enhancing iPPG Signal

When light from the projector I hits the skin, a large portion of it b is reflected directly off the skin surface without penetrating the skin and interacting with the tissue underneath. This portion of light does not contain any pulsatile information. A small portion of light penetrates the skin

(a) (b)

Figure 3: Separation of a frame into (a) Global component (b) Direct component. The direct component image shows the specular reflections from the oils on the skin surface.

surface and gets modulated by the pulsatile blood volume change waveform $p(t)$. The α term represents the strength of the modulated light. $q(t)$ is the quantization noise.

$$y(t) = I(\alpha * p(t) + b) + q(t) \tag{2}$$

One way to quantify the improvement of signal quality is by looking at the ratio of the subsurface reflectance α term and the surface reflection b. Three reflectance terms need to be considered. When the horizontally polarized light from the projector reaches the skin surface, it can (1) enter the skin and reach the pulsed blood beneath, (2) bounce off the surface of the skin into the camera, or (3) it could enter the top layer of the skin without reaching the blood vessels.

For quantifying the method using cross-polarizers, let us define the horizontal component of the polarized light as having a superscript h and the vertical component of light as having a superscript v. In the first reflectance case where the light enters the skin (the subsurface scattering case α) the horizontally polarized light will lose its polarization state and will be randomly polarized into $\alpha^v/2 + \alpha^h/2$. In the second case, the light does not enter the skin and will retain its predominantly horizontally polarized state (specular surface reflectance b_S^h). The third case is when the light enters only the top portion of the skin that does not contain pulsatile blood (sub-surface reflectance that does not have blood interactions $b_{NS}^v/2 + b_{NS}^h/2$). When the light reaches the vertical polarizer at the camera side, horizontally polarized components will be eliminated to yield the following:

$$\frac{\alpha}{b_S + b_{NS}} = \frac{\alpha^v/2}{b_{NS}^v/2} = \frac{\alpha^v}{b_{NS}^v} \tag{3}$$

The light that enters the camera sensor does not contain the horizontally polarized specular surface reflectance term which improves the signal quality.

The structured illumination method also improves the signal quality by removing the surface reflectance term. The direct component which is the specular surface reflection term b_S acts as an all-pass filter, letting all high and low spatial frequencies to be reflected back to the camera. The sub-surface scattering allows only the low spatial frequencies to be reflected back. The subsurface to surface reflectance ratio can be described as the following. We note the high-frequency signals as having a superscript w and low-frequency signals as having a superscript l.

$$\frac{\alpha}{b_S + b_{NS}} = \frac{\alpha^l}{b_S^w + b_S^l + b_{NS}^l} = \frac{\alpha^l}{b_S^l + b_{NS}^l} \tag{4}$$

The global-separation eliminates the specular surface reflectance term, thereby increasing the ratio.

4. Video Processing and PPG Extraction

4.1. Structured illumination Direct-Global separation

The global component was extracted from each frame of the video using single-image separation as proposed by Nayar et. al [29]. Although there are ways to separate the global component with other patterns mentioned earlier, these patterns require more than one image, which in turn requires us to use a camera with a higher frame rate, an extension that we plan to study in the future. For 30fps, the time difference between frames is 33.3 ms which results in an ambiguity of ± 16.6 ms [7]. A projector was used to project high-frequency binary green and dark pattern of 2-pixel width on the face. For each pixel, we assigned it the maximum or minimum label if its brightness is the maximum or minimum within the 9x1 window around it. Then the brightness intensity values around those peaks and valleys were interpolated to obtain the full resolution L_{max} and L_{min} image. Single-image global separation was performed on each frame of the videos. The separated global and direct components are shown in figure 5. The direct component shows the specular reflections on the face. The supraorbital ridge (the area between the eyebrows), the cheek region near the nose, and the eyelids show more specularities from oils on the skin surface. After all the frames were global-component-separated, the videos were passed into the PPG estimation algorithm to estimate the PPG waveform.

4.2. PPG algorithm

The PPG estimation algorithm was used to measure the blood volume pulse, as proposed in [7] and is briefly explained as follows. For every frame, the green channel was used, because it has been proven in [10] that the absorption spectra of the prominent blood chromophores are maximum

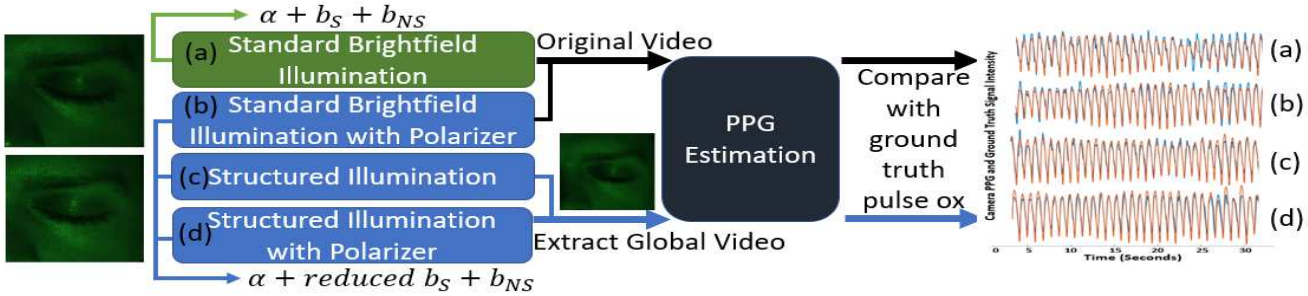


Figure 4: Videos are obtained with different illumination. The structured light videos were separated into the global-component video before passing through the PPG estimation algorithm to yield estimated PPG waveforms. The standard brightfield illumination and the videos using cross-polarizers were passed into the PPG estimation algorithm without post-processing. The α, b_S and b_{NS} terms are in reference to equations 3 and 4. b_S is the specular surface reflection term and the b_{NS} is the non-specular surface reflection term.

in the passband range of the green filters used in color cameras. Thus green channels from each frame were divided into grids of 20x20 pixels and the grids confining the face region are the Region of Interests (ROIs). These ROIs are then tracked across frames with the help of KLT Motion Tracker to reduce any effects from motion artifacts. We then averaged the pixel intensities in each of the ROIs to reduce the effect of camera sensor noise and the reflected signal intensity $y(r,t)$ is obtained as:

$$y_r(t) = I_r(\alpha_r * p(t) + b_r) + q_r(t) \tag{5}$$

where $y_r(t)$ is the pixel intensity at r^{th} ROI, b_r is the surface reflectance from the skin surface, $\alpha_r p(t)$ is the pulsatile PPG signal and q_r is the additive noise due to sensor noise, changes in illumination or motion artifacts.

Each of these signals was then filtered with a bandpass filter having cutoff frequency of (0.5Hz-5Hz). This removes additional noise which lies beyond the specified frequency range. The PPG signal estimate is then given by computing the weighted average of the signals from all ROIs

$$\hat{p}(t) = \sum_{i=1}^{N} G(i)\hat{y}(i,t) \tag{6}$$

The weights $G(r,t)$ are computed based on the idea of maximal ratio diversity [34]. Based on the assumption that the pulsatile signals from the ROIs are locally coherent, these weights are computed as a goodness metric $G(r)$ given by:

$$G(r) = \frac{\int_{PR-bw}^{PR+bw} Y(r,f)}{\int_{0.5}^{5} Y(r,f) - \int_{PR-bw}^{PR+bw}} \tag{7}$$

where $Y(r,f)$ is the power spectral density of the filtered signal $\hat{y}$ from r^{th} ROI, and PR is the pulse rate. This estimates the power of the desired signal of interest around the heart rate region. In our application, we have taken the bandwidth bw to be 0.2.

5. Methodology

The main goal of the experiment is to quantify the performance of a conventionally recorded video with a global-separated video. To evaluate how well a globally separated video performs we analyzed a video taken through a polarizer, compared it with a video that had been global-separated with structured illumination, and a combined method of using a polarizer and structured illumination at the same time.

5.1. Data acquisition

We carried out experiments on 28 human subjects (17 male, 11 female) with different skin tones from light, pale white, to dark brown. 9 subjects from light, 9 from medium, and 10 from dark skin tone categories were analyzed. The participants were asked to face the camera and be static for the duration of the video. A chin rest was provided so that the subjects could rest on it to stay as motionless as possible for two minutes. All experiments performed in this research were approved by the Rice University institute review board (Protocal number: IRB-FY2018-120 , Approval date: October 10th, 2017).

5.2. Experimental setup and data acquisition protocol

The experimental setup was organized as illustrated in Figure 3. Experiments were performed when the subject was at rest and the chin was rested on a chin rest. We used a Blackfly USB 3.0 BFLY-U3-23S6C color camera operated at 30 frames per second with a resolution of 1920x1200 and 8 bits per pixel. An Epson V355 LCD projector was used as the illumination source for all of the scenarios. The light intensities reaching the camera sensor was consistent for all

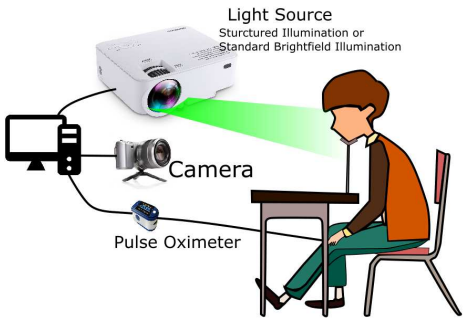


Figure 5: Setup for imaging PPG signals acquisition from face. The subject is resting comfortably facing the camera as the projector illuminates the face. A pulse oximeter was attached to the finger.

videos at 150 lux. We also used a contact pulse oximeter BIOPAC MP150 to record contact based PPG signals for comparison. The same illumination, camera, and ground truth pulse oximeter were used for all participants.

For the first set of experiment, we recorded a plain video of the subject's face at 30 fps. Standard green brightfield illumination was projected onto the face.

For the second set of experiments using structured illumination, we used a stripe pattern of repeating black and green lines that are each 2 pixels wide to project onto the face of the participants.

For the third set of experiments, we used a polarizer (Thor Labs) in front of the camera and used the standard green brightfield illumination to record videos.

In the fourth set of experiments, we used both polarizer and the structured illumination to record videos. In all these experiments, contact PPG signal was synchronously recorded from a pulse oximeter attached to the index finger.

5.3. Performance Metric

We used a signal-to-noise ratio (SNR) of the estimate to quantify the performance for all the 4 sets of experiments, and the SNR is calculated as according to the equations as follows

$$\begin{aligned} n(t) &= \hat{p}(t) - \frac{\langle \hat{p}(t), z(t) \rangle}{\langle z(t), z(t) \rangle} z(t) \\ s(t) &= \frac{\langle \hat{p}(t), z(t) \rangle}{\langle z(t), z(t) \rangle} z(t) \\ SNR &= \frac{||s(t)||^2}{||n(t)||^2} \end{aligned} \tag{8}$$

where $\hat{p}(t)$ is the extracted PPG signal, $n(t)$ is the noise, $s(t)$ is the signal of interest coherent with the ground truth signal, and $z(t)$ is the PPG signal from the contact pulse oximeter. It has been assumed that the noise contained in pulse oximeter signal is negligible compared to that obtained from a camera-based system.

6. Results

We categorized the subjects into different skin tone categories: light, medium, and dark. The average SNR for each of these categories and for each illumination are shown in the table. The standard brightfield illumination scenario had a lower SNR than the other three scenarios that remove surface reflections from the scene.
In the light skin-tone category, the global separated video using structured light performed better with an SNR improvement of an average of 1.66dB for 9 subjects (6 males, 3 females). The video using the polarizer had an improvement of 1.24dB in SNR. The structured illumination and polarizer combination improved the SNR by 1.95dB. In the light skin-tone category, the combined method using structured illumination and polarizer performed best out of the three global-separated techniques.

Tone	SBI (dB)	SI (dB)	Pol (dB)	SI + Pol (dB)
Light	3.28	4.95	4.53	5.23
Medium	3.08	4.57	4.70	5.08
Dark	2.57	3.68	4.17	4.34

Table 1: Standard brightfield illumination (SBI), Structured illumination (SI), Polarizer (Pol), and the Structured illumination and Polarizer combined method. The combined method performs the best.

In the medium skin-tone category, 9 subjects were analyzed (7 males, 2 females). The average SNR improved 1.49dB by using global-separation with structured illumination. The polarizer method improved the SNR by 1.62 dB and the structured illumination and polarizer combined method improved the SNR the most by 2dB. In the dark skin-tone category, 10 subjects were analyzed (6 males, 4 females). The SNR improvement using structured illumination was 1.11 dB. Polarizers increased the SNR by 1.60dB and the combined structured light and polarizer increased the dB by 1.76dB.

The standard brightfield illumination without global separation performed the worst. The video taken with the polarizer and the video taken with structured illumination improved the SNR. The combined method performed the best for the light, medium, and dark skin tones. In all illumination cases, the videos of light and medium skin tones performed better than that of dark skin tones.

While the average improvement is computed over the whole exposed region, the improvement in the SNR of specific skin regions (with high specular component) ranged from 3.48 dB to more than 5.63 dB. The area between the eyebrows, eyelid, and the cheek area near the nose are areas

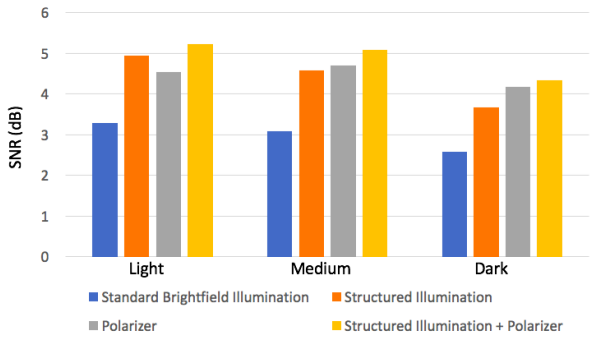


Figure 6: Three different categories of skin tone are represented. Four methods were analyzed shown from left to right: Standard brightfield illumination (SBI), Structured illumination (SI), Polarizer (Pol), and the Structured illumination and Polarizer combined method. The SNR increases with global separation and the combined structured illumination and polarizer method performs the best compared to the other methods.

with high secular reflections where this large local improvement can have a significant impact.

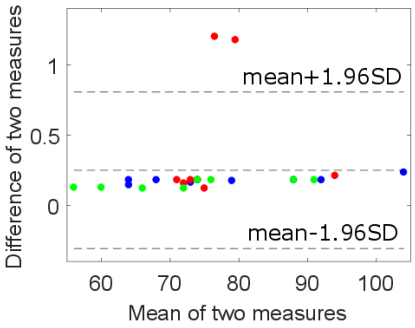
Cases:	SBI (bpm)	SI (bpm)	Pol (bpm)	SI + Pol (bpm)
HR RMSE :	0.3735	0.33	0.33	0.1680

Table 2: Global separated methods perform better for heart rate estimation. The combined sturctured illumination and polarizer method shows the lowest RMSE.

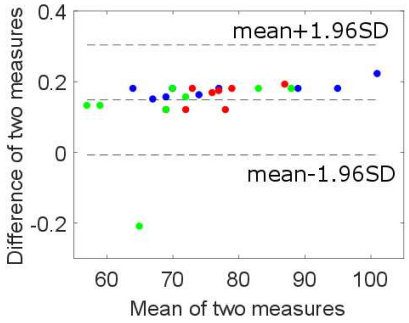
The pulse rate was also extracted from the estimated PPG signal estimates and was compared to that obtained from the contact PPG signal. Figure 7 shows the agreement of the ground truth pulse rate to the estimated pulse rate from the PPG signal for two separate cases- with standard brightfield illumination and second, with structured illumination and polarizer combined. For the standard brightfield illumination, the mean bias is 0.25 bpm with 95%limit of agreement being from -0.308 to 0.8057. By using a polarizer and structured illumination combined, the estimated error decreases to a bias of 0.1486 bpm with 95 limit of agreement between - 0.007 bpm and 0.304 bpm. The root means square error (RMSE) was also calculated for all the four sets of experiments and is listed in Table 2. The use of structured illumination and polarizer combined consistently performs better than using a simple brightfield illumination.

7. Conclusion

We have presented an improvement in the remote camera-based estimation of vital signs using direct global separation. We explored two techniques using cross-polarization and structured illumination for direct-global separation. Four experiments were performed for each of the 28 subjects: a standard brightfield illumination used for reference, structured illumination, cross-polarized, and the combine structured illumination and polarizer. PPG estimation was used to extract the waveforms for which we calculated the SNR. All of the global-separated results improved the SNR. The combined polarizer and structured illumination improve the SNR most.



(a)



(b)

Figure 7: (a) PR estimation using Simple Brightfield Illumination (b) PR estimation using Structured Illumination and Polarizer combined. The y-axis shows the difference of the ground truth and the estimated PPG signal. The x-axis is the mean of the ground truth and PPG. The skin tone is color-coded and the fair, medium and dark skin tone corresponds to the blue, red and green respectively.

8. Acknowledgements

We thank Akash Kumar Maity for helping with the writing of the paper. We also thank all of the subjects that participated in the data collection. This work was partially supported by NSF CAREER Award 1652633, NSF Expeditions Award 1730574 and NIH grant 5R01DK113269-02.

References

[1] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, p. 10762, may 2010.

[2] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 2878–2886, oct 2013.

[3] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, p. 21434, dec 2008.

[4] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Amplitude-selective filtering for remote-PPG," *Biomedical Optics Express*, vol. 8, p. 1965, feb 2017.

[5] G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiological Measurement*, vol. 35, pp. 1913–1926, aug 2014.

[6] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiology Reviews*, vol. 8, pp. 14–25, jun 2012.

[7] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express*, vol. 6, p. 1565, apr 2015.

[8] C. Zong and R. Jafari, "Robust heart rate estimation using wrist-based PPG signals in the presence of intense physical activities," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, aug 2015.

[9] M. Kumar, J. Suliburk, A. Veeraraghavan, and A. Sabharwal, "PulseCam: High-resolution blood perfusion imaging using a camera and a pulse oximeter," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, aug 2016.

[10] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 463–477, mar 2016.

[11] F. P. Wieringa, F. Mastik, and A. F. W. van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: A first step toward "SpO2 camera" technology," *Annals of Biomedical Engineering*, vol. 33, pp. 1034–1041, aug 2005.

[12] S. Hu, V. Azorin-Peris, and J. Zheng, "Opto-physiological modeling applied to photoplethysmographic cardiovascular assessment," *Journal of Healthcare Engineering*, vol. 4, pp. 505–528, dec 2013.

[13] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological Measurement*, vol. 35, pp. 807–831, mar 2014.

[14] B. D. Holton, K. Mannapperuma, P. J. Lesniewski, and J. C. Thomas, "Signal recovery in imaging photoplethysmography," *Physiological Measurement*, vol. 34, pp. 1499–1511, oct 2013.

[15] Y. Sun, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of Biomedical Optics*, vol. 16, p. 077010, jul 2011.

[16] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, aug 2012.

[17] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 7–11, jan 2011.

[18] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 1445–1457, may 2012.

[19] A. Osman, J. Turcot, and R. E. Kaliouby, "Supervised learning approach to remote heart rate estimation from facial videos," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, IEEE, may 2015.

[20] Y. Hsu, Y.-L. Lin, and W. Hsu, "Learning-based heart rate detection from remote photoplethysmography features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, may 2014.

[21] L. Wolff and T. Boult, "Constraining object features using a polarization reflectance model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 635–657, jul 1991.

[22] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH 00*, ACM Press, 2000.

[23] A. A. Kamshilin, E. Nippolainen, I. S. Sidorov, P. V. Vasilev, N. P. Erofeev, N. P. Podolian, and R. V. Romashko, "A new look at the essence of the imaging photoplethysmography," *Scientific Reports*, vol. 5, may 2015.

[24] S. G. Demos and R. R. Alfano, "Optical polarization imaging," *Applied Optics*, vol. 36, p. 150, jan 1997.

[25] S. L. Jacques and K. Lee, "titlepolarized video imaging of skin/title," in *Lasers in Surgery: Advanced Characterization, Therapeutics, and Systems VIII* (R. R. Anderson, K. E. Bartels, L. S. Bass, C. G. Garrett, K. W. Gregory, H. Lui, R. S. Malek, A. P. Perlmutter, L. Reinisch, P. J. Smalley, L. P. Tate, S. L. Thomsen, and G. M. Watson, eds.), SPIE, jul 1998.

[26] S. L. Jacques, J. C. Ramella-Roman, and K. Lee, "Imaging skin pathology with polarized light," *Journal of Biomedical Optics*, vol. 7, no. 3, p. 329, 2002.

[27] I. S. Sidorov, M. A. Volynsky, and A. A. Kamshilin, "Influence of polarization filtration on the information readout from pulsating blood vessels," *Biomedical Optics Express*, vol. 7, p. 2469, jun 2016.

[28] A. Trumpp, P. L. Bauer, S. Rasche, H. Malberg, and S. Zaunseder, "The value of polarization in camera-based photoplethysmography," *Biomedical Optics Express*, vol. 8, p. 2822, may 2017.

[29] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast separation of direct and global components of a scene using high frequency illumination," *ACM Trans. Graph.*, vol. 25, pp. 935–944, July 2006.

[30] S. Achar, S. T. Nuske, and S. G. Narasimhan, "Compensating for motion during direct-global separation," in *2013 IEEE International Conference on Computer Vision*, IEEE, dec 2013.

[31] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus," *International Journal of Computer Vision*, vol. 102, pp. 33–55, Mar 2013.

[32] D. J. Cuccia, F. Bevilacqua, A. J. Durkin, F. R. Ayers, and B. J. Tromberg, "Quantitation and mapping of tissue optical properties using modulated imaging," *Journal of Biomedical Optics*, vol. 14, no. 2, p. 024012, 2009.

[33] M. OToole, J. Mather, and K. N. Kutulakos, "3d shape and indirect appearance by structured light transport," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 1298–1312, jul 2016.

[34] D. Brennan, "Linear diversity combining techniques," *Proceedings of the IRE*, vol. 47, pp. 1075–1102, jun 1959.