SparsePPG: Towards Driver Monitoring Using Camera-Based Vital Signs Estimation in Near-Infrared

Ewa Magdalena Nowara*†, Tim K. Marks*, Hassan Mansour*, Ashok Veeraraghavan†

*Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA
†Rice University, Houston, TX

{emn3,vashok}@rice.edu, {tmarks,mansour}@merl.com

Abstract

Camera-based measurement of the heartbeat signal from minute changes in the appearance of a person's skin is known as remote photoplethysmography (rPPG). Methods for rPPG have improved considerably in recent years, making possible its integration into applications such as telemedicine. Driver monitoring using in-car cameras is another potential application of this emerging technology. Unfortunately, there are several challenges unique to the driver monitoring context that must be overcome. First, there are drastic illumination changes on the driver's face, both during the day (as sun filters in and out of overhead trees, etc.) and at night (from streetlamps and oncoming headlights), which current rPPG algorithms cannot account for. We argue that these variations are significantly reduced by narrow-bandwidth near-infrared (NIR) active illumination at 940 *nm, with matching bandpass filter on the camera. Second, the amount of motion during driving is significant. We perform a preliminary analysis of the motion magnitude and argue that any in-car solution must provide better robustness to motion artifacts. Third, low signal-to-noise ratio (SNR) and false peaks due to motion have the potential to confound the rPPG signal. To address these challenges, we develop a novel rPPG signal tracking and denoising algorithm (sparsePPG) based on Robust Principal Components Analysis and sparse frequency spectrum estimation. We release a new dataset of face videos collected simultaneously in RGB and NIR. We demonstrate that in each of these frequency ranges, our new method performs as well as or better than current state-of-the-art rPPG algorithms. Overall, our preliminary study indicates that while driver vital signs monitoring using cameras is promising, much work needs to be done in terms of improving robustness to motion artifacts before it becomes practical.*

1. Introduction

Vital signs such as the heartbeat waveform offer a way to continuously monitor the health or alertness of a person. In controlled scenarios such as hospitals, contact devices such as pulse oximeters and electrocardiographs (ECG) are the de-facto standard tools used for vital signs measurement. These contact-based devices provide accurate, robust measurements that are clinically relevant.

In several emerging applications, however, 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.

In the past decade, improvements in camera-based remote photoplethysmography (rPPG) technology [25, 17, 6, 29, 33, 7, 32, 31, 14, 27, 16] 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 this paper, we focus on one emerging application for non-contact vital signs estimation, driver monitoring, and study the various challenges facing rPPG technology in this scenario.

1.1. Driver Monitoring Using Remote Vital Signs

Driver fatigue and distraction are the leading causes of car accidents [13]. Being able to detect when a driver is not paying attention to the road or is too sleepy to continue driving could potentially lead to preventing these car accidents. Additionally, if a driver goes into a cardiac arrest or other serious and sudden health event that would impede their ability to drive they may pose a danger to others on the road. Therefore, it is desirable to detect such events before they occur and alert the driver to prevent an accident.

Changes in HR and HRV can provide insights into a person's health or psychological well being. For exam-

ple, changes in HRV have been linked to cognitive stress [22, 15, 19] and driver fatigue [24, 1]. Being able to measure vital signs continuously and unobtrusively while driving would provide a way to infer the driver's health and mental status.

1.2. Challenges and Opportunities

Seamless integration of camera-based monitoring of driver HR and HRV would be extremely useful to detect changes in driver alertness and prevent accidents. Unfortunately, the application of rPPG to driver monitoring presents several unique challenges that must be addressed head-on. We now outline three major challenges presented by in-car rPPG and explain how our proposed method addresses each of these challenges.

1.2.1 Challenge 1: Drastic Illumination Changes

During driving, illumination on the driver's face can change dramatically (Figure 1). During the day, sunlight is filtered by trees, clouds, and buildings before reaching the driver's face. As the car moves, this direct illumination changes dramatically in both magnitude and spacial extent.

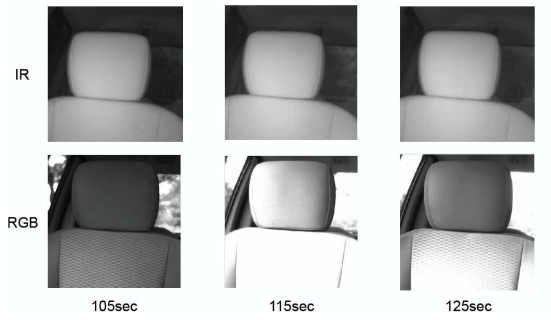

Figure 1. Example of light variations in the car during driving in IR and RGB. Narrowband IR is more robust to outside light variations.

At night, overhead streetlamps and headlights of approaching cars cause large-intensity, spatially non-uniform changes in illumination. These illumination changes are so dramatic and omnipresent that algorithmic approaches to mitigate these illumination variations are not practical.

Instead, we argue that active in-car illumination, in a narrow spectral band in which sunlight and streetlamp spectral energy are both minimal, is the most effective means to combat this challenge. Shown in Figure 2 is the spectral energy content of sunlight that reaches the Earth's surface. Due to the increased absorption near 940 nm by water in the atmosphere, sunlight at the surface has much less energy around that frequency [23]. The light output by streetlamps and vehicle headlights is typically in the visible spectrum, with very little power at infrared frequencies. Using an active illumination source at 940 nm ensures that much of the illumination changes due to environmental ambient illumination are filtered away. Further, since it is beyond the visible range, humans do not perceive this light source and thus are not distracted by its presence. Moreover, the narrower the bandwidth of the light source used in the active illumination, the narrower the bandpass filter on the camera can be, which further rejects changes due to ambient illumination. In our experiments, we used an LED source and camera bandpass filters with 10 nm bandwidth, but laser diode sources that have a 2 nm bandwidth could be used to further improve ambient light rejection.

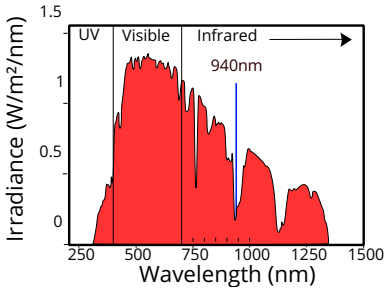

Figure 2. Spectrum of sunlight at the Earth's surface [5]. The absorption by water in the atmosphere causes a notch at 940 nm.

1.2.2 Challenge 2: Low Signal-to-Noise Ratio (SNR)

Independent of the application, one of the principal challenges for camera-based vital signs estimation is the low ratio of signal to background/noise in the raw measurements. Even in the visible spectrum, the intensity changes due to blood flow are extremely small compared to the absolute pixel intensity values. With the additional restriction of using NIR frequencies, the already weak pulsatile signal content is further minimized. Measuring rPPG in NIR is especially difficult because hemoglobin absorption is significantly lower in the infrared compared to its peak absorption in the green range of the light spectrum [29, 30, 11, 21]. The sensitivity of standard camera sensors is also decreased in the infrared range. Hence, the strength of the rPPG signal in NIR will be much lower than in the visible spectrum, as shown in Figure 3. For these reasons, the rPPG signal observed by our 940 nm system will be significantly weaker than would be observed using a regular RGB camera and broadband ambient indoor illumination. This challenging operational scenario necessitates an algorithm that is robust to noise sources.

We propose a novel robust algorithm called SparsePPG that tracks the rPPG signal over time in presence of high noise. Denoising with SparsePPG is illustrated in Figure 4. We adaptively select good facial regions and denoise their estimated rPPG signals by relying on the fact that the pulsatile signal should be sparse in the frequency domain and low-rank across facial regions. Unlike many existing methods, which require multiple wavelengths of light to achieve robustness [6, 25, 28], our algorithm is able to achieve high

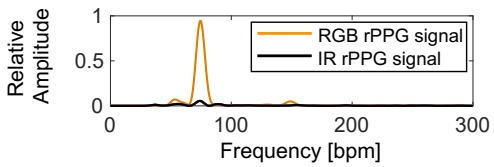

Figure 3. Comparison of rPPG signal frequency spectrum in IR and RGB. The rPPG signal is IR (blue) is about 10 times weaker than in RGB (orange).

accuracy in time-varying average heart rate estimation using a single-channel image with narrow-band illumination.

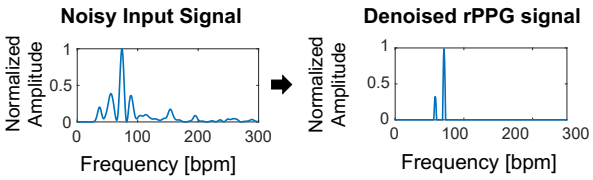

Figure 4. Example of SparsePPG algorithm denoising an rPPG signal to provide a clean rPPG estimate.

1.2.3 Challenge 3: Large Motion

Measuring rPPG while driving is additionally challenging because of large motion present in the car. The driver has to constantly turn their head to look for on-coming traffic. The motion of the car itself also causes the head of the driver to move from side to side. Lastly, the car motion causes the camera and lights to vibrate slightlty, leading to erroneous rPPG estimates.

We propose a novel algorithm that is based on sparsity in the Fourier domain that has the potential to handle large magnitude motions. There are two key ideas that work together to help handle large motions. First, we use face alignment (face landmark localization) and facial region tracking to explicitly account for and compensate for as much of the motion as possible. Second, we use a sparse Fourier representation that has the potential to separate the true peaks of the rPPG signal from false peaks due to motion.

1.3. Contributions

The main contributions of this paper are:

1. We study the applicability of camera-based vital signs measurement for driver monitoring and quantify the major challenges facing this application.
2. We develop a novel system that includes active illumination at 940 nm and show that it is feasible to accurately estimate heart rate using narrow-band NIR light around a single frequency.
3. We propose a novel algorithm, SparsePPG, to estimate the heartbeat signal in a high noise rPPG setting. The algorithm leverages the joint sparsity in the frequency spectra of rPPG signals across facial regions. The joint sparsity characterizes the common heartbeat signal in the rPPG data. Motion tracking and Robust Principal Components Analysis (RPCA) are also used to suppress a large portion of the noise and reject non-pulsatile peaks in the frequency spectrum.
4. We release the first public dataset containing simultaneous NIR and RGB videos captured in the lab with ground-truth pulse oximeter recordings.
5. We achieve high accuracy in time-varying average heart rate estimation, and show that for both NIR and RGB data, our proposed method matches or exceeds the performance of state-of-the-art rPPG algorithms.

2. Related Work

Camera-based rPPG methods have reached a high level of accuracy for RGB video recordings in bright, controlled illumination, achieving robustness to some motion and variation in skin tones. In RGB, linear components analyses of the three color channels (R, G, B) have been used to separate the heart rate signal from noise. Poh et al. performed blind source separation, applying independent component analysis (ICA) on the three color channels to separate the heart rate signal from noise [25]. Later, Lewandowska et al. used principal components analysis (PCA) on RGB channels to estimate heart rate [17].

It has also been shown that rPPG can be obtained in the presence of subject motion by using a combination of multiple wavelengths of light. The chrominance features method (CHROM) of de Haan et al. [6] is based on the observation that whereas the rPPG signal is strongest in the green channel [6, 29], the noise caused by motion is similar in all three channels. Most methods that are designed to be robust to motion require multiple camera wavelengths in order to separate the motion-related noise from the rPPG signal [6, 33, 7, 32, 31]. Estep et al. alternatively used multiple cameras to overcome large motion [8].

Blackford et al. showed the feasibility of obtaining a long-distance rPPG signal in natural sunlight outdoors [3]. The only source of light variation was the naturally changing cloud cover, and subjects were stationary during the recording. However, this work did not address scenarios that have low light levels or rapidly varying illumination, which remain a challenge for rPPG in a moving vehicle and in other outdoor situations. One way to account for uncontrolled lighting is to divide by the stationary (slowly varying) portion of the signal, as suggested in [6]. This normalization addresses slowly varying illumination, but it is not sufficient to account for illumination that varies rapidly during the time window.

The rPPG signals obtained in NIR video recording are more noisy than in visible light [30], a possible reason why less work has been done using NIR images for rPPG measurement. He et al. used broadband NIR light to show the feasibility of obtaining a rPPG signal beyond the visible

spectrum [10]. Using broadband NIR light (versus narrowband) allows more light to be captured by the camera, which improves the SNR of the estimated rPPG signals. Thus broadband NIR can enable vital signs monitoring indoors at night, but it is susceptible to lighting variations that occur at any frequencies in the near-infrared spectrum. Narrowband NIR light, in contrast, has the potential to be robust to such variations.

Van Gastel et al. simultaneously recorded images using multiple narrowband wavelengths of NIR light to achieve motion robustness in NIR [28]. While this was effective for motion robustness, it was not intended to handle time-varying lighting in NIR. Robustness to lighting variations requires narrow-band NIR filters at the camera (matched to the light source frequencies), which for multiple bands requires either multiple cameras, or a multi-band filter, which can be cost prohibitive. In order to achieve robustness to outside lighting variations with a simple, cost-effective physical setup, it is advantageous to use a single narrow frequency band, to ensure that the majority of the light recorded by the camera comes from the designated illuminators rather than from uncontrolled outside light. Robust, cost-effective measurement of vital signs in varying illumination requires algorithms that can use single-channel images without the need for multiple light frequencies.

Kumar et al. showed that by rejecting noisy facial regions, and using the signal-to-noise ratio (SNR) of the remaining regions as weights, accurate heart rate estimates can be obtained using only the green color channel in presence of slight motion [14]. Based on a similar intuition that facial regions with poor-quality signals should be identified and rejected, Tulyakov et al. applied self-adaptive matrix factorization to the chrominance feature signals from several facial regions [27]. Lam et al. used blind source separation with multiple facial regions to achieve robustness [16].

The method we propose uses a similar idea of rejecting facial regions with poor signal quality and denoising the remaining regions. Our method successfully uses a single channel of narrow-band NIR light, which is effective at rejecting outside lighting variations.

3. SparsePPG

In this section, we describe SparsePPG, our proposed model for rPPG signal tracking and denoising that we apply to videos recorded with a combination of NIR illumination and unknown ambient lighting. SparsePPG extracts, tracks, and denoises the rPPG signal to obtain an accurate heart rate measurement. An overview is shown in Figure 5.

3.1. The rPPG signal model

We obtain the raw rPPG signals from NIR-illuminated video of a face by averaging the pixel intensity values within N facial regions. As shown in the top right image in Figure 5, these facial regions are focused around the forehead, cheeks, and chin area. We exclude areas along the face boundary as well as the eyes, nose, and mouth, since these areas exhibit weak rPPG signals.

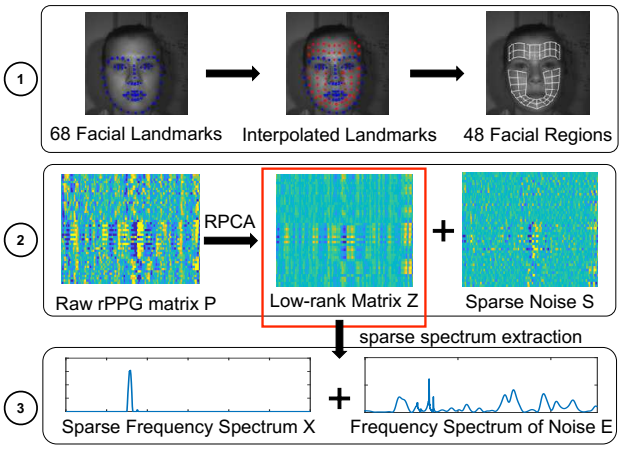

Figure 5. Overview of SparsePPG. (1) 68 landmarks are detected (blue), interpolated (red), and connected to create 48 facial regions (white). (2) Robust Principal Components Analysis (RPCA) is used to denoise the low-rank rPPG signals. (3) The signals are further denoised by finding the sparse frequency signal corresponding to the rPPG waveform.

For every facial region $j \in \{1, \ldots, N\}$, the measured mean pixel intensity $p_j(t)$ is a one-dimensional time series signal, where $t \in \{1, \ldots, T\}$ is the temporal video frame index within a temporal window of length T. We first describe the signal model based on stationary facial regions, and later describe an approach for tracking the regions over time as the subject's face moves. In the most general formulation, we may model the rPPG measurements from the N facial regions as a multichannel signal acquisition scenario, in which every facial region j provides a different channel measurement of the underlying heartbeat signal contaminated by noise. In particular, we model the measured signals $p_j(t)$ as follows:

$$p_j(t) = h_j(t) * y_j(t) + n_j(t), \tag{1}$$

where $*$ is the linear convolution operator, $y_j(t)$ denotes the heartbeat signal observed at channel j, and $h_j(t)$ and $n_j(t)$ denote the channel response function and channel noise, respectively. Since the heartbeat signal is known to be sparse in the frequency domain, we rewrite (1) in vector form as shown below

$$\mathbf{p}_j = \mathbf{h}_j * \mathbf{F}^{-1}\mathbf{x}_j + \mathbf{n}_j, \tag{2}$$

where $\mathbf{F}$ is the one-dimensional discrete Fourier transform of size T, and $\mathbf{x}_j \in \mathbb{C}^T$ denotes the sparse frequency spectrum of the heartbeat signal $\mathbf{y}_j \in \mathbb{R}^T$.

The signal model in (2) is a blind multichannel estimation problem that appears in fields such as wireless communications and sensor calibration. In particular, if $\mathbf{x}_j = \bar{\mathbf{x}}$ is

fixed across all regions j, the problem is known as the self-calibration from multiple snapshots model [18]. The recoverability of these models relies on the ability to find low-dimensional characterizations of the channel responses $\mathbf{h}_j$ and the sparse signal $\bar{\mathbf{x}}$. In this paper, we ignore the effect of the channel response functions and consider the following signal model:

$$\mathbf{p}_j = \mathbf{F}^{-1}\mathbf{x}_j + \mathbf{n}_j, \tag{3}$$

where the sparse spectrum signals $\mathbf{x}_j$ are not equal to each other, but they share the same support, i.e., the frequencies that have nonzero energy are mostly the same across all facial regions. We plan to address the more general form of the signal model in future work.

3.2. Denoising the rPPG signals

We process the rPPG data by considering sliding time windows of length T. For each time window, we stack the N rPPG signals into a $T \times N$ rPPG matrix $\mathbf{P}$. The rPPG matrix contains raw rPPG signals that are contaminated by large amounts of noise due to inaccurate motion alignment, abrupt illumination changes, and variations in the strength of the rPPG signal across regions. However, all regions should express the same periodic physiological signal caused by the cardiac cycle. Moreover, the periodicity of the underlying heartbeat signal over the duration of the temporal window induces a low-rank matrix when the noise is removed [27]. Therefore, we model the rPPG matrix $\mathbf{P}$ as the superposition of a low-rank matrix $\mathbf{Y}$ containing the heartbeat signal and a noise matrix $\mathbf{N} = \mathbf{E} + \mathbf{S}$, where $\mathbf{E}$ denotes inlier noise and $\mathbf{S}$ denotes outlier noise, such that

$$\mathbf{P} \;=\; \mathbf{Y} + \mathbf{N} \;=\; \mathbf{Y} + \mathbf{E} + \mathbf{S} \;=\; \mathbf{Z} + \mathbf{S}. \tag{4}$$

In particular, the outlier noise arises from abrupt illumination changes and region tracking errors. These generally occur over a short time duration relative to the temporal processing window and affect a small number of regions. The inlier noise characterizes regions of the face where the heartbeat signal is not the dominant driving harmonic. In this case, we wish to suppress such regions from the heartbeat signal estimation. In order to extract an estimate of $\mathbf{Y}$ from $\mathbf{P}$ and suppress outliers, we follow the robust principal component analysis (RPCA) approach of [4] and formulate the following optimization problem:

$$\min_{\mathbf{Z},\mathbf{S}} \|\mathbf{Z}\|_* + \gamma\|\mathbf{S}\|_1 \quad \text{subject to} \quad \mathbf{P} = \mathbf{Z} + \mathbf{S}, \tag{5}$$

where $\|\mathbf{Z}\|_* = \sum_k \sigma_k(\mathbf{Y})$ denotes the nuclear norm of the matrix $\mathbf{Z}$, which equals the sum of its singular values σ_k. The ℓ_1 norm of a matrix $\mathbf{S}$ is defined as $\|\mathbf{S}\|_1 = \sum_{t,j} |\mathbf{S}(t,j)|$, which equals the sum of the absolute values of its entries. The parameter γ controls the relative proportion of the signal energy that will be absorbed into the noise component $\mathbf{S}$. A smaller value of γ allows more of the signal to be considered as noise.

While many algorithms exist in the literature for solving (5), we follow the approach of Mansour et al. [20] for its speed and accuracy. In [20], the low-rank matrix $\mathbf{Z}$ is split into two factors $\mathbf{Z} = \mathbf{L}\mathbf{R}^\mathsf{T}$, where $\mathbf{L} \in \mathbb{R}^{T\times r}, \mathbf{R} \in \mathbb{R}^{N\times r}$, and $r < T, N$ is a rank estimate parameter. Notice that the RPCA model is capable of eliminating sparse outlier noise from the rPPG measurements. An illustration of denoising the rPPG signals with RPCA is shown in Part 2 of Figure 5. However, it may happen in some instances that the signal from a facial region is noisy for the entire time window. Such a noise distribution could still be modeled as low-rank, and would therefore not be removed by RPCA. We address such noise artifacts in the following section on sparse spectrum estimation.

3.3. Sparse spectrum estimation

Over a short time window, the heartbeat signal is approximately periodic, composed of a dominant frequency along with its harmonics. As a result, the frequency spectrum of a heartbeat signal should be sparse. Moreover, the same heartbeat signal drives the periodic behavior in the rPPG signals across all facial regions. Therefore, the noise-free frequency spectra $\mathbf{x}_j$ of the signals $\mathbf{y}_j$ from all regions j should have the same support.

Consider again the signal model in (4), where now we model the denoised output of RPCA as $\mathbf{z}_j = \mathbf{F}^{-1}\mathbf{x}_j + \mathbf{e}_j$ and written in matrix form below:

$$\mathbf{Z} = \;\; \mathbf{F}^{-1}\mathbf{X} + \mathbf{E} \;\; = \;\; [\mathbf{F}^{-1}\mathbf{I}] \begin{bmatrix}\mathbf{X}\\ \mathbf{E}\end{bmatrix}, \tag{6}$$

where $\mathbf{E}$ corresponds to the region level noise. Therefore, if a region is noisy, we want the entire time window (all samples) of that region to be absorbed into the matrix $\mathbf{E}$. This can be achieved by forcing complete columns of $\mathbf{E}$ to be either zero or nonzero. On the other hand, since the frequency components in $\mathbf{X}$ should be sparse and have the same support across all the regions, we want the columns of $\mathbf{X}$ to be jointly sparse, i.e., we want entire rows of $\mathbf{X}$ to be either completely zero or nonzero.

Consequently, we define the following optimization problem to compute $\mathbf{X}$ and $\mathbf{E}$ from $\mathbf{Z}$:

$$\min_{\mathbf{X},\mathbf{E}} \frac{1}{2} \left\| \mathbf{Z} - \mathbf{A} \begin{bmatrix}\mathbf{X}\\ \mathbf{E}\end{bmatrix} \right\|_2^2 + \lambda\|\mathbf{X}\|_{2,1} + \mu\|\mathbf{E}^\mathsf{T}\|_{2,1}, \tag{7}$$

where we defined $\mathbf{A} := [\mathbf{F}^{-1}\mathbf{I}]$, and the $\ell_{2,1}$ norm of a matrix X is defined as

$$\|X\|_{2,1} = \textstyle\sum_t \sqrt{\sum_j X(t,j)^2}. \tag{8}$$

The solution to the above problem can be obtained using standard iterative shrinkage/thresholding algorithms, such as FISTA [2], where the shrinkage function should be applied appropriately to the row norms of $\mathbf{X}$ and column norms of $\mathbf{E}$. Part 3 of Figure 5 shows the frequency spectra of recovered signal $\mathbf{X}$ and noise $\mathbf{E}$.

3.4. Fusion of Time Windows

Since heartbeat signals vary slowly over time, we may consider the rPPG observations as multichannel measurements from a nearly stationary process. Therefore, we process the rPPG signals using a sliding window $\mathbf{P} = \begin{bmatrix} \mathbf{P}_o \\ \mathbf{P}_n \end{bmatrix}$, where $\mathbf{P}_n$ denotes the *new* rPPG data that did not exist in the previous window, and $\mathbf{P}_o$ is the portion of the previous (*old*) window's rPPG data that is also in the current window. For better noise suppression, we construct a weighted-average time-fused window $\overline{\mathbf{P}} = \alpha\mathbf{P} + (1-\alpha)\begin{bmatrix} \widetilde{\mathbf{Y}}_o \\ \mathbf{P}_n \end{bmatrix}$, where $\widetilde{\mathbf{Y}} = \mathbf{F}^{-1}\mathbf{X}$ is the filtered output of the previous time window, and $\widetilde{\mathbf{Y}}_o$ is the portion of $\widetilde{\mathbf{Y}}$ that is also present in the current window. The time-fused window $\overline{\mathbf{P}}$ is then further denoised using the RPCA procedure, and the new sparse spectrum is estimated as described above.

3.5. Preprocessing to Reject Facial Regions

Some facial regions are physiologically known to contain better rPPG signals [14]. However, the "goodness" of these facial regions also depends on the particular video conditions, facial hair, or facial occlusions. Therefore, it is important to identify which regions are likely to contain the most noise and remove them before any processing, so that they don't affect the signal estimates. Similarly to the distancePPG method of Kumar et al. [14], we throw away a region if its signal-to-noise ratio (SNR) is below a threshold θ_{SNR} or if its maximum amplitude is above a threshold θ_{amp}. The SNR is measured as the ratio of the area under the power spectrum curve in a region surrounding the maximum peak in the frequency spectrum, divided by the area under the curve in the rest of the frequency spectrum from 30 to 300 beats per minute (bpm), as illustrated in Figure 6.

$$\text{SNR} = \frac{\int_a \mathbf{P}}{\int_b \mathbf{P} - \int_a \mathbf{P}}$$

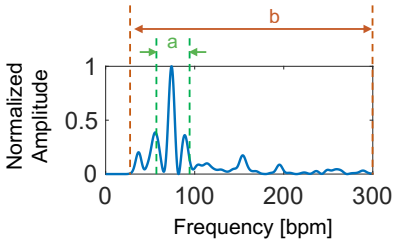

Figure 6. Definition of SNR measure: The ratio of the signal power in the region surrounding the maximum peak divided by the signal power in the rest of the range from 30 bpm to 300 bpm.

Within each time window, we may reject different facial regions. To perform fusion of time windows (see Section 3.4), we first recompose the signal $\mathbf{X}$ in the missing regions by interpolating from neighboring regions.

3.6. Implementation Details

To compute the raw rPPG signals $\mathbf{P}$, we first we detect 68 facial landmarks using the Dlib library [12], then interpolate and extrapolate the detected landmarks to a total of 145 points to include the forehead region and subdivide the face into more regions. We fix the number of facial regions for each person to be 48 We track the facial regions of interest using the Kanade-Lucas-Tomasi (KLT) tracker [26] and RANSAC algorithm [9]. In each frame, we spatially average the pixel intensities in each facial region to obtain a raw rPPG signal. We subtract the mean intensity over time of each region's signals and use a bandpass filter to restrict the signals to the frequency range [30 bpm, 300 bpm], which includes the physiological range of the cardiac signals of interest.

We use a time window of 10 seconds and an overlap between time windows, where only 10 frames come from the new time window (0.33 seconds for videos recorded at 30 frames per second (fps)). We set the rank r to 12, γ used in RCPA to 0.05, $\lambda = 0.2$, $\mu = 1$ and the threshold for fusing time windows $\alpha = 0.03$. We use SNR and amplitude threshold values $\theta_{\text{SNR}} = 0.2$ and $\theta_{\text{amp}} = 16$, set to $4\times$ average rPPG signal amplitude which was around 4 in our dataset.

To combine the denoised signals from each facial region, we take a median in each frequency bin across the regions of $\mathbf{X}$. The frequency component for which the power of the frequency spectrum is maximum corresponds to the heart rate in the given time window.

4. Experimental Evaluation and Results

In this section, we present our experimental setup and results. First, we show the feasibility of using narrow-band 940 nm illumination for rPPG measurement. Next, we demonstrate the advantages of using NIR over RGB in varying illumination. Finally, we analyze the challenges of estimating heart rate (HR) in a moving car.

4.1. HR estimation in 940 nm illumination

To test the feasibility of measuring HR in narrow-band NIR light, and to compare the performance to standard broadband RGB, we recorded videos inside the lab in controlled illumination.

We recorded videos of 12 healthy subjects (3 female, 9 male), aged 20–40, with varying skin tones (4 of the males had facial hair). The subjects were asked to sit still, but to allow for natural head motion we did not use a headrest. The raw 10-bit images were recorded with 640×640 resolution at 30 fps. The exposure for the IR camera was fixed for all participants, but the RGB camera's exposure was manually adjusted to ensure that images of people with darker skin tones were well exposed. We turned off gamma correction and set gain to zero. The videos are each about 3 minutes long. This dataset will be publicly released.

An illustration of our experimental setup is shown in Figure 7a. We simultaneously recorded videos with an RGB camera (FLIR Blackfly BFLY-U3-23S6C-C) as a benchmark, and an NIR camera (Point Grey Grasshop-

per GS3-U3-41C6NIR-C) fitted with a narrow-band 940 nm bandpass filter with 10 nm passband. We used two Bosch EX12LED-3BD-9W illuminators, with diffusers on the lights to widen the beam in order to more uniformly illuminate the face. We used ambient overhead lights to accommodate the RGB camera. We used a CMS 50D+ finger pulse oximeter to obtain a ground-truth PPG waveform.

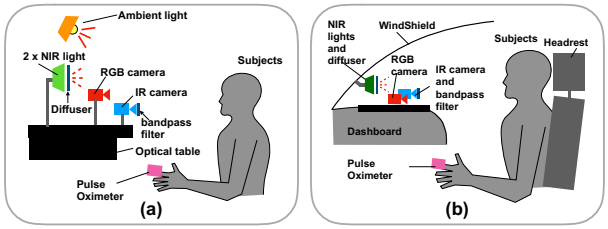

Figure 7. Schematic of the experimental setup (a) in the lab and (b) in the car, consisting of RGB and NIR cameras, NIR lights, and a finger pulse oximeter for ground-truth measurement.

4.1.1 Comparison to state-of-the-art methods

In order to compare the accuracy of our proposed algorithm, we implemented ICA [25], CHROM [6], and distancePPG [14] methods and evaluated their performance on our dataset. Because ICA and CHROM require multiple camera channels, we could only evaluate their performance on our RGB videos. RGB results using distancePPG and using our algorithm were obtained using the green channel only. For ICA and CHROM, we used spatially averaged signals from all facial regions. For distancePPG, we applied thresholds and weights on each facial region before spatially comibining them as described by Kumar et al. [14]. We increased the amplitude threshold from 4 (the value used in [14]) to 16 because, we were using 10-bit videos, so our range of values was 4 times larger than in 8-bit videos.

We used two error measures: root mean squared error (RMSE) in bpm, and the percentage of time that the HR error is less than 6 bpm (PTE6) in %. We chose an error threshold of 6 bpm because 6 bpm is the expected frequency resolution on a 10 second window assuming no zero-padding.

Table 1. HR Estimation Error on Our IR and RGB datasets

	IR		RGB	
	RMSE [bpm]	PTE6 [%]	RMSE [bpm]	PTE6 [%]
SparsePPG (Ours)	**13.6**	**79.7**	**9.6**	**88.5**
distancePPG [14]	14.7	77.7	9.7	88.4
ICA [25]	—	—	10	87.8
CHROM [6]	—	—	13.6	85.9

In Table 1, we report the results averaged over all 12 subjects. As the table shows, SparsePPG algorithm performs slightly better than the state-of-the-art camera-based rPPG methods on both the NIR and RGB videos. Notice that the performance of all methods is decreased in NIR compared to RGB due to lower rPPG signal strength in NIR.

4.2. Varying Illumination in the Lab

In the lab, we simulated the lighting variation that might occur while driving at night by manually switching the overhead lights on and off. We recorded videos simultaneously with an NIR and an RGB camera on 2 subjects. Each video lasted two minutes, divided into four 30-second intervals: 1) lights ON, 2) lights rapidly oscillating ON/OFF, 3) lights ON, and 4) lights OFF. (See the top graph in Figure 8.) NIR illuminators were constantly on. Because the tracking of the facial regions on RGB images failed in varying light and darkness, we used a headrest to minimize the amount of motion, in order to keep the positions of the facial landmarks fixed.

The light conditions, and sample results on one subject, are shown in Figure 8. The summarized results (averaged over two subjects) are shown in Table 2. To make this analysis independent of the rPPG algorithms, we used a simple spatial average of all facial regions' intensities over time as an estimate of the rPPG signals. The SNR computed for this experiment is the ratio of the area under the power spectrum around the ground truth signal frequency (instead of around the maximum peak, which was used for preprocessing in Section 3.5), divided by the area under the curve in the rest of the frequency spectrum from 30 to 300 bpm.

The results show that broadband RGB cameras are more affected than narrow-band NIR by the changing (ON/OFF) light conditions. We were not able to block all ambient light completely, so there was a small amount of light reaching the face. As a result, the RGB camera was able to accurately measure the rPPG signal in the dark, but the SNR was much lower than in bright lighting. The rPPG signals captured in NIR were accurate in all light conditions.

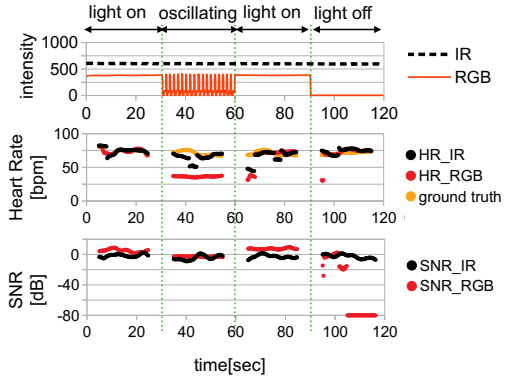

Figure 8. Varying illumination experiments. *Top:* Average pixel intensities on the face over time. *Middle:* Estimated HR in NIR (black) and RGB (red), compared to ground truth HR (orange). *Bottom:* SNR [dB] of the spatial average rPPG signal in NIR (black) and RGB (red).

Table 2. HR Estimation Error in Varying Light

	IR			RGB		
Lights	RMSE [bpm]	PTE6 [%]	SNR [dB]	RMSE [bpm]	PTE6 [%]	SNR [dB]
ON	4.1	91.7	-1.8	7.6	95.9	7.6
OFF	2.4	95.1	-5.7	7.5	95.1	-43.4
ON/OFF	2.4	95.1	-3.4	26.1	0	-12.7

4.3. Challenges in a Moving Car

We built a setup to mount RGB and NIR cameras with NIR illumination on a car dashboard (Figure 7b). We recorded a video on one subject while driving to understand the challenges for continuous HR monitoring in the car. For safety and to reduce excessive head motion, the subject was the passenger, not the driver. We found that there are two major difficulties that need to be overcome for rPPG measurements in the car: large head motion, and large illumination changes.

4.3.1 Motion in the Car

We found that even when driving at the low campus speed limit of 20 mph, the motion of the car caused the head of the passenger to move with large out-of-plane rotations so significant that the tracking of facial landmarks failed. Whenever the tracker failed, we re-detected the facial landmarks and resumed tracking. However, we found that re-detecting the facial landmarks used to define the facial regions had detrimental effects on the rPPG signals estimated from those regions. The position of the landmarks in one frame was not exactly the same as in previous frame, resulting in a slightly different intensity values in each regions and consequently corrupted rPPG signals. Moreover, the cameras and lights moved slightly during driving, causing small intensity variations.

4.3.2 Illumination Variation in the Car

To understand the challenges with varying illumination, independent of the passenger's motion, we recorded a video without the passenger in the car seat to measure how the intensity on the smooth seatback surface changes over time. We recorded videos on a sunny day, for 2 minutes while parked (stationary) and for 5 minutes while driving around campus.

We measured the average pixel intensity over time in small (20×20 pixel) regions on the left and right side of the car seat headrest. We found that the intensity changes over time were small in both NIR and RGB when the car was parked (see Figure 9). While driving, the intensity changes in RGB caused by varying outside light were large. There were also intensity changes in the NIR video recordings during driving, but they were much smaller than those in RGB. The sources of intensity changes in NIR could be caused by some outside light that is not completely blocked by the bandpass filter, or by small motion of the lights and the camera during driving.

We also found that the intensity was constantly higher on the region that was closer to the window (curve A in Figure 9) in both NIR and RGB, but the difference between the intensities of the two regions was much larger in RGB. This suggests that perhaps the NIR videos were also affected by sunlight to a small extent, even with a narrow 940 nm bandpass filter on the camera.

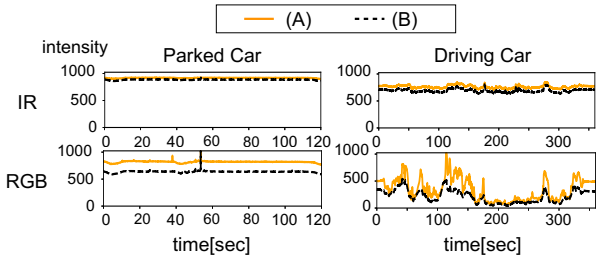

Figure 9. Intensity variations in NIR and RGB when the car was parked and while driving. The intensity variations were larger while driving than when the car was parked for both cameras, but the changes in RGB were much larger. Region A, which was closer to the window, has consistently higher intensity in both NIR and RGB.

5. Conclusions

In this paper, we tested the feasibility of using narrowband 940 nm NIR illumination to measure rPPG signals in controlled and varying light. We found that it is possible to accurately measure average heart rate in 940 nm indoors in both controlled and varying ambient light, but that the SNR is significantly decreased in NIR compared to RGB in controlled lighting conditions.

We presented SparsePPG, a new algorithm for denoising raw rPPG signals that achieves state-of-the-art accuracy in HR estimation. SparsePPG is able to achieve a high accuracy using a single camera channel by relying on the fact that rPPG signals should be sparse in frequency and low-rank with across facial regions.

We conducted an analysis of sources of error in a realistic scenario in the car. We built a set up with RGB and NIR cameras and NIR illumination, and recorded data while driving. We found that large motion and illumination changes significantly affect both RGB and NIR video recordings, making rPPG measurements very difficult. However, we found that NIR data are much less affected by these changes.

Acknowledgments

Ashok Veeraraghavan and Ewa Nowara were partially supported by NSF CAREER Award 1652633, NSF Expeditions Award 1730574 and NIH grant 5R01DK113269-02.

References

[1] H. J. Baek, H. B. Lee, J. S. Kim, J. M. Choi, K. K. Kim, and K. S. Park. Nonintrusive biological signal monitoring in a car to evaluate a drivers stress and health state. *Telemedicine and e-Health*, 15(2):182–189, 2009. 2

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. 5

[3] E. B. Blackford and J. R. Estepp. Measurements of pulse rate using long-range imaging photoplethysmography and sunlight illumination outdoors. In *Optical Diagnostics and Sensing XVII: Toward Point-of-Care Diagnostics*, volume 10072, page 100720S. International Society for Optics and Photonics, 2017. 3

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 5

[5] Commons Wikimedia. Spectrum of solar radiation (earth). https://en.wikipedia.org/wiki/Sunlight#/media/File:Solar_spectrum_en.svg, 2013. [Online; accessed 10-April-2018.] CC BY-SA 3.0. https://creativecommons.org/licenses/by-sa/3.0/. Figure was adapted to only show sunlight spectrum as seen at the Earth's surface. 2

[6] G. De Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013. 1, 2, 3, 7

[7] G. De Haan and A. Van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014. 1, 3

[8] J. R. Estepp, E. B. Blackford, and C. M. Meier. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pages 1462–1469. IEEE, 2014. 3

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Readings in computer vision*, pages 726–740. Elsevier, 1987. 6

[10] X. He, R. Goubran, and F. Knoefel. Ir night vision video-based estimation of heart and respiration rates. In *Sensors Applications Symposium (SAS), 2017 IEEE*, pages 1–5. IEEE, 2017. 4

[11] M. Huelsbusch. *An image-based functional method for opto-electronic detection of skin-perfusion*. PhD thesis, Phd thesis, RWTH Aachen dept. of EE.(in German), 2008. 2

[12] V. Kazemi and S. Josephine. One millisecond face alignment with an ensemble of regression trees. In *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*, pages 1867–1874. IEEE Computer Society, 2014. 6

[13] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, D. J. Ramsey, et al. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. 2006. 1

[14] M. Kumar, A. Veeraraghavan, and A. Sabharwal. Distanceppg: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express*, 6(5):1565–1588, 2015. 1, 4, 6, 7

[15] D. Kurian, K. Radhakrishnan, and A. A. Balakrishnan. Drowsiness detection using photoplethysmography signal. In *Advances in Computing and Communications (ICACC), 2014 Fourth International Conference on*, pages 73–76. IEEE, 2014. 2

[16] A. Lam and Y. Kuno. Robust heart rate measurement from video using select random patches. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3640–3648, 2015. 1, 4

[17] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcama non-contact method for evaluating cardiac activity. In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, pages 405–410. IEEE, 2011. 1, 3

[18] S. Ling and T. Strohmer. Self-calibration and bilinear inverse problems via linear least squares. *arXiv:1611.04196*, 2016. 5

[19] C. R. Madan, T. Harrison, and K. E. Mathewson. Noncontact measurement of emotional and physiological changes in heart rate from a webcam. *Psychophysiology*, 2017. 2

[20] H. Mansour, D. Tian, and A. Vetro. Factorized robust matrix completion. *Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, 2016. 5

[21] L. F. C. Martinez, G. Paez, and M. Strojnik. Optimal wavelength selection for noncontact reflection photoplethysmography. In *22nd Congress of the International Commission for Optics: Light for the Development of the World*, volume 8011, page 801191. International Society for Optics and Photonics, 2011. 2

[22] D. McDuff, S. Gontarek, and R. Picard. Remote measurement of cognitive stress via heart rate variability. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 2957–2960. IEEE, 2014. 2

[23] B. Park, Y. Keh, D. Lee, Y. Kim, S. Kim, K. Sung, J. Lee, D. Jang, and Y. Yoon. Outdoor operation of structured light in mobile phone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2392–2398, 2017. 2

[24] H. Park, S. Oh, and M. Hahn. Drowsy driving detection based on human pulse wave by photoplethysmography signal processing. In *Proceedings of the 3rd International Universal Communication Symposium*, pages 89–92. ACM, 2009. 2

[25] 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*, 18(10):10762–10774, 2010. 1, 2, 3, 7

[26] C. Tomasi and T. Kanade. Detection and tracking of point features. 1991. 6

[27] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, pages 2396–2404, 2016. 1, 4, 5

[28] M. van Gastel, S. Stuijk, and G. de Haan. Motion robust remote-ppg in infrared. *IEEE Transactions on Biomedical Engineering*, 62(5):1425–1433, 2015. 2, 4

[29] W. Verkruysse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008. 1, 2, 3

[30] V. Vizbara. Comparison of green, blue and infrared light in wrist and forehead photoplethysmography. *BIOMEDICAL ENGINEERING 2016*, 17(1), 2013. 2, 3

[31] W. Wang, B. Balmaekers, and G. de Haan. Quality metric for camera-based pulse rate monitoring in fitness exercise. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2430–2434. IEEE, 2016. 1, 3

[32] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Color-distortion filtering for remote photoplethysmography. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 71–78. IEEE, 2017. 1, 3

[33] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Robust heart rate from fitness videos. *Physiological measurement*, 38(6):1023, 2017. 1, 3