MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Near-Infrared Imaging Photoplethysmography During Driving

Nowara, Ewa; Marks, Tim; Mansour, Hassan; Veeraraghavan, Ashok

TR2020-161 December 16, 2020

Abstract

Imaging photoplethysmography (iPPG) could greatly improve driver safety systems by enabling capabilities ranging from identifying driver fatigue to unobtrusive early heart failure detection. Unfortunately, the driving context poses unique challenges to iPPG, including illumination and motion. First, drastic illumination variations present during driving can overwhelm the small intensity-based iPPG signals. Second, significant driver head motion during driving, as well as camera motion (e.g., vibration) make it challenging to recover iPPG signals. To address these two challenges, we present two innovations. First, we demonstrate that we can reduce most outside light variations using narrow-band near-infrared (NIR) video recordings and obtain reliable heart rate estimates. Second, we present a novel optimization algorithm, which we call AutoSparsePPG, that leverages the quasi-periodicity of iPPG signals and achieves better performance than the stateof-the art methods. In addition, we release the first publicly available driving dataset that contains both NIR and RGB video recordings of a passenger's face with simultaneous ground truth pulse oximeter recordings.

IEEE Transactions on Intelligent Transportation Systems

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Near-Infrared Imaging Photoplethysmography During Driving

Ewa M. 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*—Imaging photoplethysmography (iPPG) could greatly improve driver safety systems by enabling capabilities ranging from identifying driver fatigue to unobtrusive early heart failure detection. Unfortunately, the driving context poses unique challenges to iPPG, including illumination and motion. First, drastic illumination variations present during driving can overwhelm the small intensity-based iPPG signals. Second, significant driver head motion during driving, as well as camera motion (e.g., vibration) make it challenging to recover iPPG signals. To address these two challenges, we present two innovations. First, we demonstrate that we can reduce most outside light variations using narrow-band near-infrared (NIR) video recordings and obtain reliable heart rate estimates. Second, we present a novel optimization algorithm, which we call AutoSparsePPG, that leverages the quasi-periodicity of iPPG signals and achieves better performance than the state-of-the-art methods. In addition, we release the first publicly available driving dataset that contains both NIR and RGB video recordings of a passenger's face with simultaneous ground truth pulse oximeter recordings.**

***Index Terms*—remote photoplethysmography, imaging photoplethysmography, near-infrared, heart rate, driver monitoring**

I. INTRODUCTION

EVERY year, there are 6 million car accidents in the U.S., of which 94% are caused by human error, including distraction and fatigue [2], [3]. Furthermore, heart disease is the leading cause of death—every 40 seconds someone suffers from a heart attack in the U.S. [4]. If a heart attack happens during driving, the driver is no longer able to control the vehicle and poses an immediate threat to himself and to others present on the road. Continuous and unobtrusive vital signs measurements could prevent a large number of these accidents by early detection of fatigue [5], distraction [6], and potentially even life-threatening episodes such as heart attacks and tachycardia [7]–[13].

Over the last few years, camera-based measurement of vital signs, including heart rate (HR) [14], breathing rate [15], and heart rate variability (HRV) [15], has reached sufficient accuracy to have potential in diverse realistic applications [14], [16], [17]. These measurements of vital signs with a camera are known as imaging photoplethysmography (iPPG). They are derived from minuscule intensity variations of skin regions with each cardiac cycle, caused by varying blood volume over time. Remotely measuring vital signs with cameras could improve driver monitoring systems and be seamlessly incorporated inside the car, without requiring the user to wear a contact device.

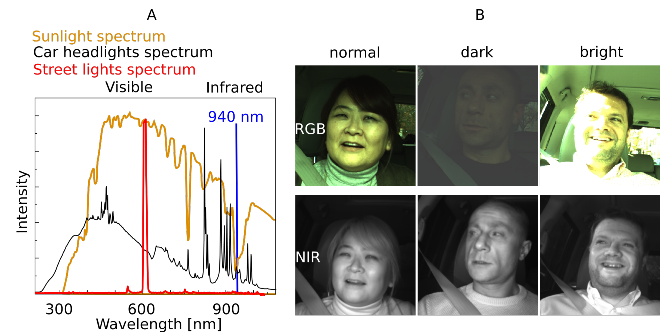


Fig. 1. A. Spectrum of ambient light sources present during driving. Most of the ambient light is reduced in NIR, especially around 940 nm [25]–[27]. B. RGB cameras are more susceptible to ambient light variations than NIR cameras.

In addition to measuring vital signs, recording a driver's face with a camera can provide information about gaze [18], head pose [19], blinking rate [20], changes in facial expression [21], and other subtle facial parameters [22]–[24], for more accurate multi-modal measurements of the driver's mental and health status.

Unfortunately, there are unique sources of noise in a moving vehicle that make most existing iPPG methods unsuitable for this application. First, the outdoor ambient light varies drastically and suddenly during driving (e.g., while driving through the shadows of buildings, trees, etc.), making it difficult to distinguish iPPG signals from other intensity variations. Second, there is significant motion of the driver's head and face due to a number of factors, such as the motion of the car, the driver looking around both within and outside the car (for oncoming traffic, looking into rear- and side-view mirrors, etc.), and the driver talking. Third, there are currently no publicly available datasets with video recordings captured during driving that have simultaneous ground truth measurements of vital signs. Therefore, it is difficult to fully understand the challenges that driving poses for iPPG measurement.

While iPPG methods using RGB color cameras are more robust to motion than iPPG using near-infrared (NIR), they fail in presence of uncontrolled ambient illumination. On the other hand, iPPG using NIR cameras (with NIR illumination that is invisible to the subject) can be robust in most illumination settings, but it is not as robust to large motion as RGB methods. However, when large motion and light variations are present, no existing methods, in either NIR or RGB, work well (See

TABLE I
ROBUSTNESS OF NIR AND RGB SYSTEMS TO DIFFERENT LIGHT AND MOTION SETTINGS.

Camera	Low Light	Varying Light	Controlled Light	Small Motion (MR-NIRP Indoor [1])	Large Motion (e.g., Driving)	Large Motion and Low / Varying Light
NIR	✓	✓	✓	✓	×	×
RGB	×	×	✓	✓	✓	×

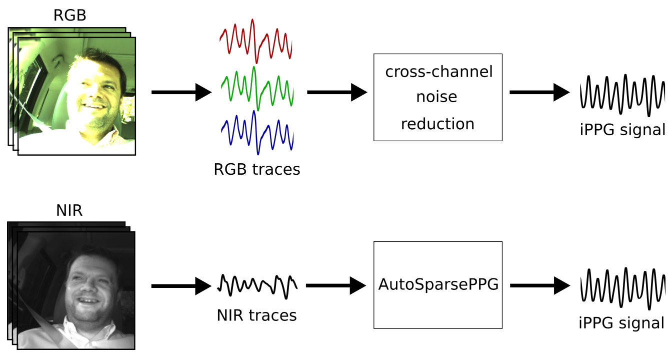
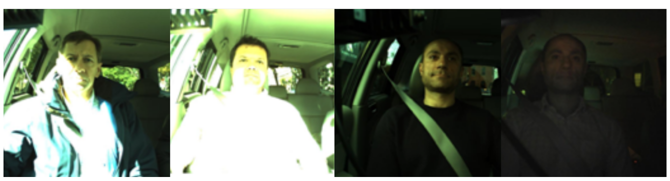


Fig. 2. The top row illustrates state-of-the-art approaches for measuring iPPG signals with RGB camera recordings, which leverage multiple camera channels to obtain motion-robust iPPG signals. However, RGB cameras are susceptible to ambient light variations. The bottom row illustrates our proposed monochromatic NIR system, which is robust to ambient light variations, and our AutoSparsePPG algorithm that is capable of robustly recovering iPPG signals in the presence of motion.



a. non-uniform b. saturated c. day d. night

Fig. 3. Examples of ambient light variations in RGB during driving.

Table I). In this work, we present a system-level solution. We leverage the robustness of narrow-band NIR to uncontrolled illumination, and we design an AutoSparsePPG algorithm to enable robustness to motion in NIR. The contributions of this paper include the following:

1) **Hardware**: We design a narrow-band NIR system, and we find an optimal wavelength range that reduces the majority of ambient light variations during driving, including sunlight (see Fig. 1).
2) **Algorithm**: We develop an iPPG algorithm robust to motion which outperforms the state-of-the-art methods in NIR recordings (see Fig. 2).
3) **Dataset**: We release the first publicly available video dataset that contains video recordings in RGB and NIR captured during driving, as well as synchronized ground truth pulse oximeter measurements.[1]

II. CHALLENGES FOR IPPG IN THE CAR

In videos, iPPG signals are detectable as minuscule amplitude variations modulating the intensity of skin pixels. Due to the weakness of the iPPG signal, existing techniques for estimating iPPG are highly susceptible to nuisance factors that affect image intensity. In order for iPPG techniques to be successfully deployed in car-related applications, the primary challenges that need to be overcome are ambient illumination variations and motion of both the car and the person in the car.

A. Ambient illumination variations

Because iPPG is a low-intensity signal, the signal-to-noise ratio needs to be amplified by signal processing techniques, such as spatial averaging and incorporating information from multiple heartbeat cycles. In most existing work, temporal averaging is performed over 5–10 seconds (about 5–10 cardiac cycles) [28]. In many applications, it is reasonable to assume that ambient illumination is constant over this time span, and that the intensity variations on a stationary subject's face are primarily due to iPPG variations. In the driving context, however, traditional algorithms that assume constant or slowly varying illumination do not work well.

There are several illumination-based challenges for iPPG in the driving context. First, during driving the amount of light falling on the driver's face can change suddenly and drastically, as sunlight is blocked and revealed by buildings or trees during the day, or as the car drives underneath streetlamps or past oncoming vehcles' headlights at night. Second, the ambient light can illuminate different facial regions from different angles and with different brightness. This results in a non-uniform pattern of light and shadow across the face (Fig. 3a), making it difficult to directly combine these facial regions to compute iPPG signals. Third, there is a high dynamic range across time and space. The driver's face may be very bright (even completely saturated) when it is in direct sunlight (Fig. 3b), but very dark either when the car is in the shadow cast by a building during the day (Fig. 3c) or at night (Fig. 3d). As a result of these high-frequency, high-amplitude spatio-temporal variations in facial illumination, traditional iPPG algorithms that operate on RGB videos fare poorly in driving applications.

B. Large motion

During driving, the car's velocity changes frequently due to the driver engaging the accelerator or the brakes, steering around turns, and traversing hills and bumps in the road. All of these changes in velocity produce involuntary motion of the driver's head. Moreover, the driver exhibits both rigid head motion (looking around for oncoming traffic, looking into the rear- or side-view mirrors, and looking at other objects inside or outside the car) and non-rigid facial motion (talking, singing, eating, or making facial expressions). See Fig. 4 for

[1]Our MR-NIRP Car Dataset may be downloaded here: https://computationalimaging.rice.edu/databases/.

a. out-of-plane rotation b. facial expressions

Fig. 4. Examples of sources of motion present during driving.

motion examples. Head motion can rapidly change the surface normals at each pixel, leading to substantial changes in image intensity that often overwhelm the minuscule iPPG signals. Additionally, the motion of the car causes some vibrations of the camera and lights used for data collection.

We compared the amount of facial motion of subjects under two categories of car motion and two categories of subject motion. The car was either parked in a garage with the engine running or driving in a city. The subjects were either asked to sit as still as possible so that any motion (e.g., due to changes in car velocity) was unintentional, or they were asked to behave as if they were driving: look out the windshield, glance at rear-view and side-view mirrors, and talk naturally. We computed the amount of facial motion within each 10-second time window as the average Euclidean distance between the positions of each detected facial landmark in two adjacent frames:

$$\frac{1}{K}\sum_{k=1}^{K-1}\sum_{t=1}^{T-1}\sqrt{(T_{t+1,k,x}-T_{t,k,x})^2+(T_{t+1,k,y}-T_{t,k,y})^2},$$

where K is the number of facial landmarks and T is the duration of the time window. We averaged the motion (measured in pixels) over all 10-second time windows and across all subjects.

The amount of involuntary motion caused only by the moving car was large ($\sim$224 pixels)—comparable to the amount of voluntary motion performed when the car was still ($\sim$216 pixels). The amount of voluntary motion was twice as large as the amount of involuntary motion, both during driving (voluntary:$\sim$415 pixels, involuntary:$\sim$224 pixels) and when the car was parked (voluntary:$\sim$216 pixels, involuntary:$\sim$114 pixels), making this a very challenging high-noise scenario. An average size of the face for these videos was 130 $\times$ 180 pixels.

C. Lack of publicly available driving datasets

As the challenges facing iPPG estimation in the car are fundamentally different from those in more stationary applications, such as video-conferencing [29], datasets that were acquired in other contexts are not useful to study iPPG estimation in the driving context. Almost all existing publicly available iPPG datasets were captured indoors, with RGB cameras and with controlled illumination. Some of these datasets have head motion, but the motion is mostly caused by facial expressions and talking, which is radically different from the head motion caused by a moving vehicle. Consequently, previous datasets are not useful for understanding the challenges for iPPG during driving. There have been few papers attempting to measure iPPG in the car using cameras [8], [30], and so far only one group is planning to publicly release their dataset [30]. Therefore, it is difficult to understand how the ambient light and motion artifacts impact the iPPG signal quality during driving, and how much more severe these artifacts are compared to those in indoor recordings.

III. RELATED WORK

A. Algorithms based on multiple color channels

Almost all state-of-the-art iPPG algorithms achieving the highest accuracy and motion robustness rely on combinations of the [R, G, B] channels. Linear combinations of the color channels can be used to separate the heart rate signal from noise [14], [31], [32]. However, the use of RGB cameras requires sufficiently bright and controlled visible light, and will not work well at night or when the ambient light is varying drastically. Van Gastel et al. used three NIR cameras, each fitted with a different narrow-band filter, to achieve robustness to both motion and light variations [33]. However, using multiple cameras can be cost-prohibitive, and image registration from multiple camera views may be challenging.

B. Algorithms applicable to monochrome recordings

When the ambient illumination is either dark or varying rapidly, as in the driving context, monochrome NIR recordings (with NIR illumination) are a cost-effective way to eliminate illumination-based noise. However, most of the state-of-the-art algorithms use three color channels (e.g., RGB) for motion robustness, so they will not work on monochrome recordings. There have been a few algorithmic solutions proposed that model the properties of the iPPG signal without relying on multiple channels. Kumar et al. showed that by identifying which facial regions have strong signals and weighing them by their SNR measure, robust heart rate estimates can be obtained using only the green color channel [28]. We proposed the SparsePPG algorithm, which leveraged the fact that iPPG signals are sparse in the frequency domain and low-rank across facial regions [1]. However, many of these methods require setting fixed optimization parameters or thresholds beforehand, making it hard to generalize them to new datasets with different cameras or lighting conditions.

C. Addressing uncontrolled illumination

Blackford et al. demonstrated the feasibility of obtaining iPPG measurements with RGB cameras outdoors with sunlight as the source of illumination [34], but the subjects were stationary and the outside light was not varying suddenly. Chen et al. [35] used broadband NIR light for night and low-light settings. However, the iPPG signals obtained using NIR are much more noisy than using visible light [36], [37]. Using broadband NIR recordings improves the signal strength compared to narrow-band NIR because it allows more light to be captured by the camera. However, broadband NIR is still susceptible to ambient light variations that may occur at NIR frequencies, especially those caused by sunlight. We demonstrate the feasibility of using very narrow-band (10 nm passband) filters with a monochrome NIR camera to achieve heart rate estimation accuracy comparable to benchmark methods that use RGB cameras.

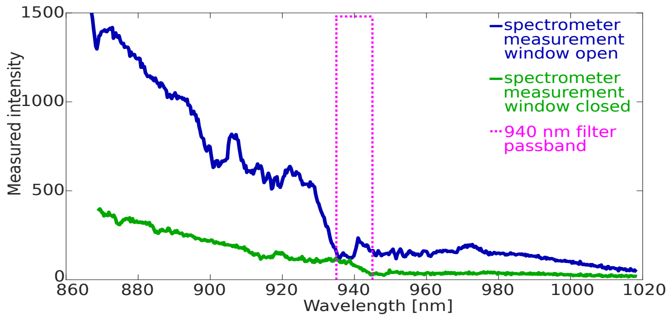


Fig. 5. Sunlight spectrum measured with the car window open (solid blue line) and closed (solid green line), and passband of the 940 $\pm$ 5 nm bandpass filter (dotted magenta line). The majority of sunlight energy that reaches the car interior is in wavelengths shorter than 930 nm. The glass in the car window also blocks a significant portion of the NIR light from the sun.

D. iPPG during driving

There have been few papers that attempted to estimate iPPG during a realistic driving scenario. Kuo et al. used spatially averaged green-channel intensities from a single region on the face to compute the average heart rate during driving, but obtained accuracy below 20% for more than half of the subjects in their dataset [8]. Wu et al. used a k-nearest neighbor classifier applied to the strongest five frequency peaks of the chrominance signal's frequency spectrum [7]. However, they only reported results for two participants. We release the first publicly available iPPG driving dataset and show that we obtain reliable results in heart rate estimation using the proposed NIR set up.

IV. Handling Illumination Noise With NIR Imaging

A. Illumination variations in visible and NIR

The sources of the ambient light in images captured during driving include sunlight, streetlamps, and headlights of other vehicles. Most modern street lighting uses low-pressure sodium lamps which predominantly emit visible light [38]. The light bulbs used in car headlights are usually halogen or xenon bulbs which also mostly operate in the visible range, though they also emit some NIR light. However, sunlight contains energy in NIR, visible, and ultraviolet wavelengths (see Figure 1).

We measured which wavelengths dominate the ambient light in the car due to sunlight by capturing measurements with a spectrometer (Ocean Optics USB2000+) through a car window that was either open or closed (see Figure 5). There is a large drop off in the energy at longer wavelengths, starting at about 930 nm. The reason for the sudden drop-off in the spectral energy of the sunlight at the Earth's surface at around 930 nm is that water in the atmosphere absorbs light in a wavelength band that includes 940 nm [39]. The other sources of illumination variations during driving, such as street lights and headlights, are predominantly in the visible spectrum and can also be reduced using an NIR filter.

B. Design choices in active NIR imaging

In this subsection, we discuss several design choices and trade-offs necessary to consider when building an active NIR illumination system for measuring iPPG.

1) Achieving Uniform Illumination: Illumination intensity across the face can be non-uniform due to the variation in the 3D directions of the normals across the face surface, due to shadows cast on the face, and due to different parts of the face being at different distances from each illuminator. To make the illumination more uniform across the face, we used two light sources, placed on each side of the face and at roughly equal distances from the head. In addition, we placed horizontal and vertical diffusers on the light sources to widen the light beams reaching the face, so that the center of the face would not be much more brightly lit than the periphery.

2) Capturing Well-Exposed Images: We would like the images of the face to be sufficiently well exposed in order to measure strong iPPG signals. However, the intensity of the illumination is inversely proportional to the square of the distance from the light source to the face. If the face is too close to the illumination, the images will be saturated and will not contain iPPG information, but as the person moves farther back from the lights, the images will become dimmer and have weaker iPPG signals. It is also important to keep the camera exposure fixed during the duration of the recording to obtain clean iPPG signals. We experimentally selected the most favorable position of the illuminators inside the car and their brightness setting to avoid capturing saturated images, while recording well-exposed images at a range of possible distances between the subject's face and the camera. We tested different distances (ranging from 7 cm to 50 cm) by having the participant sit inside the car and lean forward and backward, while the position of the camera was fixed.

3) Bandwidth, Light Efficiency, and Eye Safety: The more narrow the optical bandpass filter on the camera, the more ambient light can be rejected, reducing the amount of noise corrupting the iPPG signals during driving. However, when very narrow filters are used, the captured images become dark and the strength of the iPPG signals decreases. Therefore, using narrow-band filters requires using bright illumination matching the passband wavelength of the filter to obtain well-exposed images.

NIR light can be shined on a person's face without causing discomfort because it is invisible to the human eye, making it easy to use very bright lights. Because the NIR lights are not visible to the human eye, however, the pupillary light reflex does not narrow the pupils to limit the amount of light reaching the eyes, even when very bright NIR lights are used. Consequently, care needs to be taken to ensure that the NIR illumination is within the eye safety limits. We conducted these computations (included in Supplementary Materials), according to the OSRAM eye safety note [40].

C. Lower iPPG SNR in NIR

The iPPG signals are strongest in the green part of the light spectrum because of larger absorption of hemoglobin in that wavelength range [37]. In contrast, iPPG recordings in NIR are significantly weaker and less robust to noise than recordings captured in RGB [16], [36], [37]. Moreover, most camera sensors' sensitivity decreases in the NIR range with increasing wavelength, leading to larger camera quantization noise. Finally, monochrome recordings do not enable using redundant information in multiple camera channels for de-noising, which is commonly used for motion robustness in RGB recordings (see Fig. 2) [14], [31], [32]. In summary, while narrow-band NIR can be used to reduce the noise due to ambient light, it is at the cost of lower signal-to-noise ratio (SNR) and less robustness to motion. Therefore, one must use iPPG algorithms that can be robust to motion in the low-SNR regime of NIR.

V. Algorithms: Motion Compensation

A. Computing iPPG signals from video intensities

As blood flows through a skin region, the concentration of hemoglobin changes over time, changing the amount and color of light absorbed by the skin. When we record a video of a skin region, a camera can register those small intensity variations caused by blood flow, referred to as the iPPG signal.

The quantization noise of the camera sensor, $v_n(t)$, can be reduced by spatial averaging of groups of pixels, which is a commonly used pre-processing step for extracting iPPG signals from a video recording. Therefore, we obtain the raw iPPG signals from the video frames by spatially averaging the pixel intensities within each of $N = 48$ regions of interest on the face. We define those $N = 48$ facial regions by first detecting 68 facial landmarks using the OpenFace library [41], then interpolating and extrapolating the detected landmarks to obtain a total of 145 points that include the forehead region, as shown in part 1 of Fig. 6. We focus on regions around the forehead, cheeks, and chin area, because these regions tend to exhibit stronger iPPG signals [28]. We exclude noisy regions such as eyes, nose, mouth, the boundary of the face, and the very top of the forehead that often contains hair. For every facial region $j \in \{1, \ldots, N\}$, the raw iPPG signal $p_j(t)$ obtained from the mean intensities is a one-dimensional time series signal, where $t \in \{1, \ldots, T\}$ is the video frame index within a time window of length T frames. We stack the signals from all N facial regions into an iPPG matrix $\mathbf{P}$ of size $T \times N$. We process the iPPG signals within 10-second sliding time windows that overlap by $\frac{1}{3}$ second (10 frames overlap for our 30 frames per second (fps) videos). We used a 10-second window to process the signals, following [28], because it was short enough to accommodate the heart rate variations, but long enough to be robust to variations in noise over time.

We normalize each time window's signals by subtracting the mean intensity over time of each region's signals and then dividing by that mean. We bandpass-filter the signals to restrict them to the standard cardiac frequency range [42 bpm, 240 bpm] [32].

1) Temporal averaging of facial landmark locations: When we detect facial landmarks in each frame independently, there is a high-frequency jitter in the position of the detected landmarks, even when the face is stationary. This causes the pixels included in different small facial regions to correspond

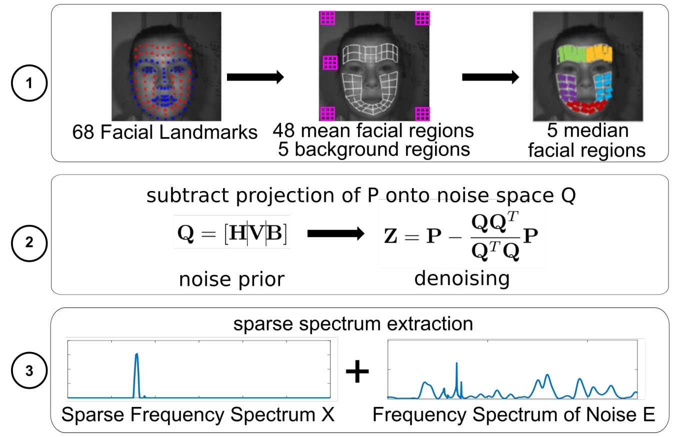


Fig. 6. Overview of our proposed AutoSparsePPG algorithm. (1) PPG signals are computed from each facial region. (2) We suppress noise components using projections onto the orthogonal complement of the motion noise ($\mathbf{H}, \mathbf{V}$) and the noise from ambient light variations $\mathbf{B}$ (computed from background regions). (3) From the partially denoised iPPG signals $\mathbf{Z}$, we then recover the quasiperiodic iPPG signal's sparse frequency spectrum ($\mathbf{X}$).

to slightly different regions on the face for each video frame, changing the average intensities over time and leading to small errors that accumulate over time and affect the iPPG performance. We found that when there is large motion and lighting variations, tracking algorithms tend to make errors that compound over time, causing the estimated positions of the facial landmarks to drift away from the correct facial locations. Instead, we estimate the position of each facial landmark in frame t by averaging the detected positions of the landmark from frame $t-5$ to $t+5$.

2) Motion robustness using median of regions: For additional robustness to small variations in facial regions' positions over time, we grouped the $N = 48$ small regions (called *mean regions*, because the signal for each small region is the mean over all of its pixels) into five larger regions with a spatial median, as shown on the right in part 1 of Fig. 6. We call the five larger regions *median regions*, because the signal for each larger region is obtained by computing for each time step the median across the signals from the small regions that make up the large region. As we detail in Supplementary Materials, using the five large median regions improves performance by as much 9% compared to only using the 48 mean regions.

3) Pre-processing by discarding noisy facial regions: Some of the facial regions may be severely corrupted by noise for a long time during driving (e.g., due to occlusions or shadows), or they may simply not contain physiologically strong iPPG (e.g., due to hair) [28]. In that case, the iPPG signals cannot be recovered from these regions, and including them in our data would corrupt the final heart rate estimates.

We assume that iPPG signals are weak and slowly varying intensity variations, so any regions that have very large energy within a short time window should be removed as likely containing noise. We remove regions with ℓ_2 norms exceeding the threshold of median$(||\mathbf{P}_t||_2) + \frac{1}{2}\sigma(||\mathbf{P}_t||_2)$, where σ is the standard deviation, computed over all five facial regions for each considered time window. The ℓ_2 norm is computed over time, and the standard deviation is computed over the five spatial regions.

B. Reducing noise using orthogonal projections

Different facial regions may be contaminated differently by noise caused by changes in ambient illumination, motion alignment errors, and facial expressions, so the noise may be high-dimensional. However, blood flows through all facial regions with approximately the same temporal profile during the cardiac cycle, so the underlying iPPG signal should be low-rank across facial regions [42]. To suppress noise that is corrupting the iPPG signal, we orthogonally project (OP) the noisy iPPG signal $\mathbf{P}$ onto the noise subspace $\mathbf{Q}$, then subtract this projected signal from $\mathbf{P}$. This is equivalent to projecting the noisy iPPG signal onto the orthogonal complement of the noise subspace.

We approximate the motion noise by summarizing the motion of the face with two time-varying 5-dimensional (5D) signals: a 5D horizontal motion signal $\mathbf{H}$, and a 5D vertical motion signal $\mathbf{V}$. We also compute a 5D time-varying background illumination signal $\mathbf{B}$, to approximate the noise caused by time-varying illumination at various locations.

To extract the 5D horizontal motion signal $\mathbf{H}$, we first measure the horizontal motion of each of the $N = 48$ small facial regions by spatially averaging the positions of the four corners of the region in each frame. We then reduce those 48 dimensions into five dimensions, one for each large region, by computing the median of the motion signals across all of the small regions that belong to that large region. The sequence of these 5D signals across all time steps in the 10-second time window is the $T \times 5$ matrix $\mathbf{H}$. The 5D vertical motion signal $\mathbf{V}$ is computed similarly.

To obtain a 5D time-varying signal $\mathbf{B}$ that represents the background ambient light intensity variation, we selected five regions in the background not containing the face area, shown in magenta in the center image of part 1 of Fig. 6. We split each of these five large background regions into small (30×30 pixel) regions, spatially average the intensity values within each small region, and take the median of the resulting nine values to obtain a single value for the large region. We do this for each of the five large background regions in each frame, resulting in a $T \times 5$ matrix $\mathbf{B}$. Finally, we concatenate these three noise sources into one noise signal matrix $\mathbf{Q} = [\mathbf{H} \,|\, \mathbf{V} \,|\, \mathbf{B}]$ of dimensions $T \times 15$. We orthogonally project the noisy iPPG signal matrix $\mathbf{P}$ onto the noise subspace $\mathbf{Q}$ and subtract that projection from the iPPG signal matrix $\mathbf{P}$, to obtain the OP-denoised signal $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{P} - \frac{\mathbf{Q}\mathbf{Q}^T}{\mathbf{Q}^T\mathbf{Q}}\mathbf{P}. \quad (1)$$

C. AutoSparsePPG: adaptive sparse spectrum estimation

1) Sparse spectrum estimation: iPPG signals are quasiperiodic, which means that they have slowly varying frequency. Over a short time window, they can be approximated as periodic signals with a dominant frequency and its harmonics. Thus, we can model the iPPG signals as sparse in the frequency domain. All facial regions containing iPPG should have the same sparse frequency spectrum and the same support of the frequency coefficients, corresponding to the underlying noise-free heartbeat signal. We model the OP-denoised signal $\mathbf{Z}$ as a sum of two components: the desired iPPG signal $\mathbf{Y}$, whose frequency spectrum, $\mathbf{X}$, has only a few non-zero coefficients; and the inlier noise, $\mathbf{E}$, that was not removed by OP:

$$\mathbf{Z} = \mathbf{Y} + \mathbf{E} = \mathbf{F}^{-1}\mathbf{X} + \mathbf{E}, \quad (2)$$

where $\mathbf{F}^{-1}$ is the inverse Fourier transform.

We want the columns of $\mathbf{X}$ to be jointly sparse to ensure that the frequency components are sparse and all regions have the same support, resulting in entire rows of $\mathbf{X}$ being either zero or non-zero. Conversely, we want to be able to remove facial regions which are noisy in the whole time window. We additionally make sure that the energy in the remaining facial regions is not very large, because the iPPG signals are very weak signals and large amplitudes likely correspond to noise. To do so, we force the entire columns to be either zero or non-zero by formulating this problem following the SparsePPG approach [1] with a mixed $\ell_{2,1}$ norm regularization:

$$\min_{\mathbf{X},\mathbf{E}} \frac{1}{2} \left\|\mathbf{Z} - \mathbf{F}^{-1}\mathbf{X} - \mathbf{E}\right\|_F^2 + \lambda \left(\|\mathbf{X}\|_{2,1} + \mu\|\mathbf{E}^\mathsf{T}\|_{2,1}\right), \quad (3)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, and the $\ell_{2,1}$ norm of a matrix X is defined as

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

The $\ell_{2,1}$ regularization is applied such that the ℓ_2 norm of the columns of $\mathbf{X}$ (facial regions) is followed by an ℓ_1 norm along the rows (frequency coefficients) to ensure sparsity within the computed column norms. Conversely, the ℓ_2 norm of the rows of $\mathbf{E}$ (time dimension) is followed by an ℓ_1 norm across the columns (facial regions) to sum up the row norms and ensure sparsity across the facial regions.

2) Adaptive regularization parameter selection: The choice of regularization parameters, λ and μ, has a significant impact on the performance of heart rate estimation. Changing either of these parameters can lead to as much as a 30% difference in HR estimation accuracy. Moreover, very different parameter values are optimal for different videos.

We propose the AutoSparsePPG algorithm, which automatically selects the parameter λ adaptively based on the data. Following the work of Van den Berg et al. for solving sparse optimization problems with least squares constraints [43], we can rewrite (3) as:

$$\begin{array}{ll} \min\limits_{\mathbf{X},\mathbf{E}} & ||\mathbf{Z} - \mathbf{F}^{-1}\mathbf{X} - \mathbf{E}||_F^2 \\ \text{subject to} & \|\mathbf{X}\|_{2,1} + \mu\|\mathbf{E}^\mathsf{T}\|_{2,1} < \tau, \end{array} \quad (4)$$

where τ is defined as:

$$\tau = \tau_0 + \frac{||\mathbf{Z} - \mathbf{F}^{-1}\mathbf{X} - \mathbf{E}||_F^2}{\max([||\nabla_\mathbf{X}||_{2,\infty}, \mu||\nabla_\mathbf{E}||_{2,\infty}])}.$$

Here, $\tau_0 = ||\mathbf{X}||_{2,1} + \mu||\mathbf{E}^\mathsf{T}||_{2,1}$ for some initial $\mathbf{X}$ and $\mathbf{E}$, and $\nabla_\mathbf{X}$ and $\nabla_\mathbf{E}$ are the gradients of $||\mathbf{Z} - \mathbf{F}^{-1}\mathbf{X} - \mathbf{E}||_F^2$ with respect to $\mathbf{X}$ and $\mathbf{E}$, respectively.

The parameter λ is initialized to λ_0:

$$\lambda_0 = \frac{||\mathbf{Z}||_F}{\sqrt{\mathrm{card}(\mathbf{X})\mathrm{card}(\mathbf{E})}}$$

where card is the cardinality (the number of the elements of the matrix). Then in each iteration, λ is updated using Newton's root finding method applied to the equation

$$\|\mathbf{X}\|_{2,1} + \mu\|\mathbf{E}^\mathsf{T}\|_{2,1} = \tau.$$

Consequently, we use the following update rule to modifying λ in order to satisfy the τ constraint:

$$\lambda_{k+1} = \max\left(0,\ \lambda_k + \frac{\|\mathbf{X}_{k+1}\|_{2,1} + \mu\|\mathbf{E}_{k+1}^T\|_{2,1} - \tau}{\beta(\|\mathbf{X}_{k+1}\|_{2,0} + \mu\|\mathbf{E}_{k+1}^T\|_{2,0})}\right),$$

where β is a step size parameter, $\|\mathbf{X}_{k+1}\|_{2,0}$ computes the number of nonzero column-norms of $\mathbf{X}_{k+1}$, and $\mathbf{X}$ and $\mathbf{E}$ are initialized with zeros. A detailed description of the AutoSparsePPG framework is presented in Algorithms 1 and 2. Algorithm 2 details our algorithm for the $\text{proj}_{2,1}$ step (Step 7) of Algorithm 1. Please see [43] for details about the convergence and stability of Newton's root finding method.

Algorithm 1 AutoSparsePPG algorithm for solving (4)

input: $\mathbf{Z}, \mathbf{X}^0, \mathbf{E}^0, \alpha$
set: $\tau_0 = \|\mathbf{X}^0\|_{2,1} + \mu\|\mathbf{E}^{0\mathsf{T}}\|_{2,1}$
1: $\nabla_{\mathbf{X}^0} \leftarrow \mathbf{X}^0 - \mathbf{F}\left(\mathbf{Z} - \mathbf{E}^0\right)$
2: $\nabla_{\mathbf{E}^0} \leftarrow \mathbf{E}^0 + \mathbf{X}^0 - \mathbf{Z}$
3: $\tau \leftarrow \tau_0 + \frac{||\mathbf{Z}-\mathbf{F}^{-1}\mathbf{X}_0-\mathbf{E}_0||_F^2}{\max([||\nabla_{\mathbf{X}_0}||_{2,\infty}, \mu||\nabla_{\mathbf{E}_0}||_{2,\infty}])}$
4: **for** $k \leftarrow 1$ to K **do**
5: $\quad \tilde{\mathbf{X}}^k \leftarrow \mathbf{X}^{k-1} - \alpha\nabla_{\mathbf{X}^{k-1}}$
6: $\quad \tilde{\mathbf{E}}^k \leftarrow \mathbf{E}^{k-1} - \alpha\nabla_{\mathbf{E}^{k-1}}$
7: $\quad (\mathbf{X}^k, \mathbf{E}^k) \leftarrow \text{proj}_{2,1}(\tilde{\mathbf{X}}^k, \tilde{\mathbf{E}}^k, \tau)$
8: $\quad \nabla_{\mathbf{X}^k} \leftarrow \mathbf{X}^k - \mathbf{F}\left(\mathbf{Z} - \mathbf{E}^k\right)$
9: $\quad \nabla_{\mathbf{E}^k} \leftarrow \mathbf{E}^k + \mathbf{X}^k - \mathbf{Z}$
return: $\mathbf{X}^K, \mathbf{E}^K$

Algorithm 2 $\text{proj}_{2,1}$: constrained $\ell_{2,1}$ projector

input: $\tilde{\mathbf{X}}, \tilde{\mathbf{E}}, \tau, \alpha.$
set: $\lambda \leftarrow 0, \mathbf{X} \leftarrow \tilde{\mathbf{X}}, \mathbf{E} \leftarrow \tilde{\mathbf{E}}$
define: $\boldsymbol{R}(X, E) := \|X\|_{2,1} + \mu\|E^\mathsf{T}\|_{2,1}$
Compute row and column norms
1: $\mathbf{X_r} \leftarrow [\|\mathbf{X}(1,:)\|_2, \ldots \|\mathbf{X}(T,:)\|_2]^\mathsf{T}$
2: $\mathbf{E_c} \leftarrow [\|\mathbf{E}(:,1)\|_2, \ldots \|\mathbf{E}(:,J)\|_2]$
3: **while** $\boldsymbol{R}(\mathbf{X}, \mathbf{E}) > \tau$ **do**
Apply soft-thresholding
4: $\quad \mathbf{X} \leftarrow \frac{\tilde{\mathbf{X}}}{\mathbf{X_r}} \odot \max\{\mathbf{X_r} - \alpha\lambda; 0\}$
5: $\quad \mathbf{E} \leftarrow \frac{\tilde{\mathbf{E}}}{\mathbf{E_c}} \odot \max\{\mathbf{E_c} - \mu\alpha\lambda; 0\}$
Compute row and column norms
6: $\quad \mathbf{X_r} \leftarrow [\|\mathbf{X}(1,:)\|_2, \ldots \|\mathbf{X}(T,:)\|_2]^\mathsf{T}$
7: $\quad \mathbf{E_c} \leftarrow [\|\mathbf{E}(:,1)\|_2, \ldots \|\mathbf{E}(:,J)\|_2]$
Update λ
8: $\quad g \leftarrow -\|\mathbf{X_r}\|_0 - \mu\|\mathbf{E_c}\|_0$
9: $\quad \lambda \leftarrow \max\left\{0, \lambda - \frac{\boldsymbol{R}(\mathbf{X},\mathbf{E})-\tau}{\alpha g}\right\}$
return: $\mathbf{X}, \mathbf{E}$

To combine the denoised signals from each facial region, we take a median in each frequency bin across the regions of $\mathbf{X}$. A median is more robust to outliers than a mean when some of the facial regions are corrupted by noise. We found that when we instead used a mean in each frequency bin, the results were often erroneous in the presence of noise. The frequency component for which the power of the frequency spectrum is maximum is the heart rate output by our algorithm for the given time window. Part 3 of Fig. 6 illustrates an example of the sparse frequency spectrum of the underlying iPPG signal recovered from noisy video data.

D. Fusion of time windows

The human heart rate varies slowly over time, so the iPPG signals from multiple facial regions can be approximated to be a stationary process within a short time window. By using the information from previous time windows, we can improve the iPPG denoising and remove a lot of abrupt changes caused by noise. We process the iPPG signals using sliding time windows. For each time window, the signal to be processed is a weighted average of two sources: the previous time window's already processed and denoised data, and the current time window's noisy data that has not yet been processed.

This weighted average is defined as follows:

$$\overline{\mathbf{P}} = \alpha \begin{bmatrix} \mathbf{P}_o \\ \mathbf{P}_n \end{bmatrix} + (1-\alpha) \begin{bmatrix} \widetilde{\mathbf{Y}}_o \\ \mathbf{P}_n \end{bmatrix}. \tag{5}$$

Here, $\begin{bmatrix} \mathbf{P}_o \\ \mathbf{P}_n \end{bmatrix}$ represents the unprocessed, noisy data from the current time window. $\mathbf{P}_o$ denotes the data from the portion of the current time window that overlaps with the previous (*old*) time window, while $\mathbf{P}_n$ denotes the data from the *new* portion of the current time window (the portion that does not overlap with the previous time window). Note that the old data, $\mathbf{P}_o$, were already processed (denoised) in the previous time step; the processed, denoised version of $\mathbf{P}_o$ (which was output at the previous time step) is denoted $\widetilde{\mathbf{Y}}_o$. The parameter α controls how much we weigh the previous window's results. The smaller the value of α, the more we take into account the previous time window's estimate.

As part of the pre-processing within each time window, we may have rejected a different number of facial regions, resulting in different dimensions of the input iPPG signals in the consecutive time windows. Therefore, after processing each time window, we first recompose the signal in the missing regions by linearly interpolating from neighboring regions in order to use the described weighted time window fusion. We selected all optimization coefficients that gave the best performance on our data by using a leave-one-subject-out cross-validation.

VI. MR-NIRP Car Dataset

In this section we present the experimental conditions and the setup used to collect our new dataset, the MERL-Rice Near-Infrared Pulse (MR-NIRP) Car Dataset.

A. Data collection conditions

To decouple the effects of motion and ambient light variations on the quality of the iPPG signals, we recorded videos in

Fig. 7. Sample video frames from our new MR-NIRP Car Dataset, in NIR (left image of each pair) and RGB (right image of each pair).

the car in two different driving conditions: inside the garage, and driving in the city. Inside the garage, the engine was running but the car was parked. During the driving scenario, we drove around the block in the city, where we often had to stop at traffic lights. Sudden stopping, accelerating, and turning introduced additional motion artifacts, making it more difficult to recover iPPG signals than it would be while driving at a constant speed on a highway. In each of the two driving conditions, we recorded data with two different head motion conditions. In the first (*minimal head motion*), we asked the participants to sit as still as possible. In the second (*additional head motion*), we asked the participants to talk naturally and to look through the windshield and in the side and rear-view mirrors to simulate the amount of motion that would be present during natural driving.

We collected data on 18 healthy subjects[2] aged 25–60 years with varying skin tones. One of the subjects was recorded twice during driving, once during the day and once at night. Therefore there are 19 recordings with city driving, and 18 recordings for condition in the garage. Of the 18 subjects, two subjects were female, and five subjects had facial hair. We recorded four videos at night and 14 during the day. Of those 14 videos, eight were recorded in sunny weather and six in overcast conditions. Examples of captured images in NIR and RGB are shown in Fig. 7. None of our participants wore glasses during the data collection. However, the presence of glasses should not significantly affect the performance of any algorithms evaluated on our dataset since the eye region is excluded, as shown in Fig. 6 part 1. All of the NIR recordings were included, but we had to exclude the RGB recordings of two subjects during city driving and one subject in the garage because the video frames were so dark that facial landmarks were not detected. We had the subjects sit in the passenger seat (the subjects did not control the car) during recording, for two reasons: for safety; and to reduce the amount of hand motion in order to avoid corrupting the pulse oximeter signals. This was important, because we found that even small involuntary motions of the hand significantly affected the recorded pulse oximeter (ground truth) signals.

B. Experimental setup

We mounted the NIR (Point Grey Grasshopper GS3-U3-41C6NIR-C) and the RGB (FLIR Grasshopper3 GS3-PGE-23S6C-C) cameras next to each other on the dashboard in front of the subject. The lenses we used had a focal length of 8 mm for the NIR camera and 4.5 mm for the RGB camera. The NIR camera was fitted with a 940 nm hard-coated optical density bandpass filter from Edmund Optics with a 10 nm passband. We also compared the performance with a 975 nm bandpass filter with a 50 nm passband and "dark frame subtraction" to further reduce ambient light, however we found there was not a significant improvement in the results (see the Supplementary Materials). We used four Bosch EX12LED-3BD-9W illuminators, two on each side of the subject's face. Each illuminator was fitted with a 95-degree diffuser in the vertical direction and an 80-degree diffuser in the horizontal direction, to broaden the beam of light and to make sure that the illumination of the face was relatively uniform. We used ambient illumination for the RGB camera. We obtained a ground-truth PPG waveform using a CMS 50D+ finger pulse oximeter recorded at 60 fps.

We recorded 10-bit raw images with 640×640 resolution at 30 fps, with no gamma correction and with fixed exposure that was set at the beginning of the video capture to make sure the face was well lit. When the images were well exposed, we always set the gain to zero, and when it was very dark, we increased the gain until the face region was sufficiently bright. All the recordings captured inside the garage were 2 minutes long; the recordings captured during driving ranged from 2–5 minutes in duration, depending on how long it took us to drive around the block.

VII. Results

A. Compared benchmark algorithms

We compared the performance of our proposed AutoSparsePPG algorithm to five state-of-the-art iPPG methods: SparsePPG [1], DistancePPG [28], ICA [14], CHROM [31] and POS [32] (detailed in the Supplementary Materials). Since ICA, CHROM, and POS require multiple camera channels, they cannot be applied to NIR recordings. To evaluate the single-channel (monochromatic) methods AutoSparsePPG, SparsePPG, and DistancePPG on the RGB recordings, we used only the green channel. Comparisons of variations of AutoSparsePPG, including the use of 48 mean facial regions rather than 5 median facial regions, are in the Supplementary Material. We evaluated all compared methods on the same videos, using the same pre-processing and the same time window parameters as our proposed method.

To evaluate the performance of the compared algorithms, we use two error measures: (i) root mean squared error (RMSE) computed between the ground truth and estimated heart rate

[2]The study was approved by MERL's Institutional Review Board, and all participants signed an informed consent form for the use and public release of their data.

TABLE II
HR ESTIMATION ERRORS ON MR-NIRP CAR DATASET ("MINIMAL" HEAD MOTION)

	Driving Day		Driving Night		Driving All		Garage	
	NIR	**RGB**	**NIR**	**RGB**	**NIR**	**RGB**	**NIR**	**RGB**
	PTE6 [%] (higher is better)							
AutoSparsePPG	**60.0 ± 6.0**	33.1 ± 3.2	**64.7 ± 12.0**	19.7 ± 7.7	**61.0 ± 5.2**	31.5 ± 3.1	**81.9 ± 5.9**	**91.1 ± 1.9**
SparsePPG [1]	18.3 ± 4.2	22.1 ± 3.6	14.1 ± 2.4	17.2 ± 6.3	17.4 ± 3.4	21.5 ± 3.2	35.6 ± 6.8	53.6 ± 9.7
DistancePPG [28]	25.5 ± 2.8	18 ± 2.6	21.2 ± 2.5	14.1 ± 6.9	24.6 ± 2.3	17.6 ± 2.4	37.4 ± 4.0	74.7 ± 5.5
ICA [14]	N/A	**57.3 ± 3.6**	N/A	**32.8 ± 1.8**	N/A	**54.4 ± 3.8**	N/A	83.3 ± 5.4
CHROM [31]	N/A	54.2 ± 4	N/A	30.2 ± 0.6	N/A	51.4 ± 4	N/A	82.6 ± 4.9
POS [32]	N/A	23.7 ± 6.1	N/A	13.4 ± 13.4	N/A	22.5 ± 5.5	N/A	52.9 ± 11.3
	RMSE [bpm] (lower is better)							
AutoSparsePPG	**11.8 ± 2.0**	$>$ 15 bpm	**11.2 ± 4.4**	$>$ 15 bpm	**11.6 ± 1.8**	$>$ 15 bpm	**5.1 ± 1.4**	**2.9 ± 0.4**
SparsePPG [1]	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm
DistancePPG [28]	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	$>$ 15 bpm	8.2 ± 1.8
ICA [14]	N/A	**9.7 ± 1.2**	N/A	**12.9 ± 1.3**	N/A	**10.1 ± 1.1**	N/A	5.3 ± 1.6
CHROM [31]	N/A	$>$ 15 bpm	N/A	$>$ 15 bpm	N/A	$>$ 15 bpm	N/A	10.5 ± 3.6
POS [32]	N/A	$>$ 15 bpm	N/A	$>$ 15 bpm	N/A	$>$ 15 bpm	N/A	$>$ 15 bpm

TABLE III
HR ESTIMATION ERRORS ON MR-NIRP INDOOR DATASET [1]

	Stationary		Motion	
	PTE6 [%] (higher is better)			
	NIR	RGB	NIR	RGB
AutoSparsePPG	**93.8 ± 2.3**	93.7 ± 2.4	**61.6 ± 8.4**	70.6 ± 8.2
SparsePPG [1]	69.5 ± 16.7	89.9 ± 10.1	41.7 ± 15.0	79.8 ± 5.4
DistancePPG [28]	72.5 ± 6.6	91.0 ± 4.7	38.0 ± 4.1	57.0 ± 8.4
ICA [14]	N/A	**98.3 ± 1.0**	N/A	88.6 ± 3.5
CHROM [31]	N/A	98.0 ± 1.1	N/A	**90.5 ± 3.8**
POS [32]	N/A	97.4 ± 1.8	N/A	89.1 ± 5.5
	RMSE [bpm] (lower is better)			
	NIR	RGB	NIR	RGB
AutoSparsePPG	**2.0 ± 0.5**	1.9 ± 0.5	**12.7 ± 2.6**	11.0 ± 3.8
SparsePPG [1]	22.5 ± 12.3	8.2 ± 7.7	24.0 ± 10.4	5.3 ± 2.1
DistancePPG [28]	8.5 ± 2	2.6 ± 1.1	18.5 ± 3.2	16.0 ± 3.7
ICA [14]	N/A	0.8 ± 0.2	N/A	**2.5 ± 0.7**
CHROM [31]	N/A	**0.8 ± 0.2**	N/A	3.2 ± 1.1
POS [32]	N/A	3.0 ± 2.2	N/A	4.9 ± 2.6

(HR) over all ten-second time windows, and (ii) percentage of the time that the HR error is less than 6 bpm (PTE6). We chose an error threshold of 6 bpm because it is the expected frequency resolution on a ten-second window. Unlike RMSE, which can be strongly impacted by large outliers (e.g., an estimated heart rate that is extremely incorrect for a short period of time), PTE6 can be thought of as roughly measuring the percent of time that the estimated heart rate is correct vs. incorrect.

B. MR-NIRP Car Dataset

The results on our new MR-NIRP Car dataset for the *minimal head motion* condition are summarized in Table II. There were often large and sudden movements of the head caused by the motion of the car, even though the participant was trying to sit still. When RMSE errors were larger than 15 bpm, we have replaced those results with RMSE "$>$ 15 bpm" to indicate that heart rate was estimated incorrectly and that the iPPG signal was not recovered well. Our proposed AutoSparsePPG method significantly outperforms all state-of-the-art methods on NIR videos. On RGB videos, AutoSparsePPG outperforms the state-of-the-art single-channel methods (SparsePPG and DistancePPG) both while driving and while parked in the garage. On RGB videos, AutoSparsePPG (which uses only the green camera channel) outperforms even the methods that use three camera channels (ICA, CHROM, and POS) while parked in the garage, but it does not do as well as them while driving. This is because driving induces significant head motion, which three-channel methods are better able to suppress. Due to the large amount of head motion in this condition, methods that use three camera channels often perform better than the single-channel methods on RGB videos.

Despite having only one channel, our NIR method performs slightly better (has higher PTE6) than the best 3-channel RGB method during daytime driving, and performs much better than the best RGB method during night driving. In summary, we achieve the following improvements with our proposed NIR hardware and AutoSparsePPG algorithm:

- Despite the head motion that is present during driving, our NIR setup with our AutoSparsePPG algorithm outperforms the state-of-the-art RGB algorithms in all driving conditions, achieving higher PTE6 by 6.6% on average (Driving All conditions).
- During daytime driving, our system slightly outperforms the best RGB method, with PTE6 higher by 2.7% (Driving Day).
- Our NIR method achieves the most significant improvements over RGB methods during night driving when it is dark, increasing PTE6 by 31.9% (Driving Night).
- Our proposed NIR setup and AutoSparsePPG algorithm are robust in all lighting conditions and partially robust to motion, whereas RGB methods fail when the illumination is too low (Driving Night).

While parked in the garage, there was enough light for RGB methods and not much lighting variation; hence, accuracy is high for both NIR and RGB (PTE6 $>$ 80%). However, our NIR method performs a bit worse than RGB in this setting, probably because the iPPG signal is stronger in visible frequencies than in NIR.

The videos collected with *additional head motion* during driving in fact had very large motion caused by the subject looking around and talking combined with motion due to the car accelerating, stopping, starting, and turning. Consequently, most methods performed very poorly on these driving videos in both NIR and RGB (with PTE6 $<$ 30% for most methods).

These results are summarized in Table I of the Supplementary Material.

C. MR-NIRP Indoor Dataset

We additionally compared the performance of AutoSparsePPG to several state-of-the-art methods on the publicly available MERL-Rice Near-Infrared Pulse (MR-NIRP) Indoor Dataset [1], which has simultaneous RGB and NIR recordings captured in an indoor setting and was the only publicly available dataset that had narrow-band NIR videos with motion. Since before this paper there were no publicly available driving datasets with ground truth physiological signals, MR-NIRP Indoor is the most similar existing dataset to our new MR-NIRP Car dataset. The MR-NIRP Indoor dataset contains videos recorded of subjects seated in a lab performing two tasks: a *stationary* task and a *motion* task. In the stationary task, subjects were asked to sit still for 3 minutes. In the motion task, subjects were asked to count out loud from zero to ten and perform random slight head motion.

On NIR videos, our proposed AutoSparsePPG outperformed all other methods in both the stationary and motion tasks (Table III). The results of AutoSparsePPG on stationary NIR videos are close to the performance of the best RGB methods on stationary RGB videos, demonstrating that NIR recordings can be nearly as robust as RGB when the motion is not very large. Most methods performed very well on the stationary task in RGB recordings because the data are clean and there are not many sources of noise. In the presence of motion, the three-channel methods ICA and CHROM performed best on RGB videos. Finally, the results of AutoSparsePPG on NIR videos are similar to its results on RGB videos (especially in the stationary task), demonstrating that our proposed algorithm is able to achieve robustness to motion and noise even in the more noisy NIR videos.

VIII. Discussion

Our experiments demonstrate that by using narrow-band NIR light sources and filter, our proposed AutoSparsePPG algorithm achieves good heart rate estimation performance that is robust to ambient light variations and low light settings, when there is not too much motion. In the presence of significant motion, however, multi-channel RGB methods are more robust. On the other hand, while three-channel RGB methods can be motion robust, they are easily corrupted by ambient light variations. Furthermore, since we cannot shine visible light on a person's face without causing discomfort and dangerous driving conditions, it is not feasible to use RGB in low light settings.

One way to achieve robustness to both ambient light and motion might be to use multiple NIR cameras to enable linear combinations of multiple NIR channels, similar to algorithms designed for RGB recordings (e.g., ICA, CHROM, POS). Alternatively, RGB and NIR cameras could be used jointly, to leverage RGB motion robustness when the lighting variations are not large, and to leverage the robustness of NIR to uncontrolled lighting when the ambient light is varying or is too dark for RGB. In both of these cases, using multiple NIR cameras can be expensive, and errors in registering images from multiple cameras may also adversely affect iPPG signals. Therefore, the most promising future avenue may be to use a single NIR camera but to develop more motion-robust algorithms.

On average, there is 7.5 to 10 BPM difference in average HR between drowsy and alert states [9], [44], so HR error less than that may be required for driving applications. We achieve the required accuracy when the car is parked, and we are close during driving but the accuracy needs to be improved by 3–4 BPM. Clinically approved gold standard contact devices have RMSE errors in average HR on the order of 3 bpm. State-of-the-art camera-based methods use an already relaxed error standard of 6 bpm. However, in a very challenging driving scenario a larger average RMSE might be acceptable if time windows that have large errors can be detected and ignored. We do not expect any method to work all the time in a very challenging driving scenario, but if we could identify time windows that have unreliable HR estimates, then the system making decisions based on these measurements could discount them.

IX. Conclusion

The presented work is the first detailed study of the sources of noise for iPPG during driving. We have identified and analyzed the unique challenges for iPPG technology posed by a realistic driving scenario, and we presented hardware and algorithmic solutions to these challenges. First, we showed that the variations in uncontrolled ambient light affecting RGB recordings during driving can be significantly reduced with a narrow-band NIR hardware set up.

Second, we showed that a degree of motion robustness can be achieved in monochrome NIR recordings with the proposed AutoSparsePPG algorithm, despite the significantly lower SNR of iPPG signals in NIR compared to the visible range. AutoSparsePPG outperformed the state-of-the-art methods that do not require multiple camera channels. While the proposed NIR set up can reduce a lot of light variations, it is not as motion robust as methods that leverage multiple RGB channels. However, while methods using multiple camera channels (e.g., ICA) and RGB recordings sometimes performed better in presence of motion, our proposed AutoSparsePPG with the NIR set up was the only method that performed consistently well in all conditions—in presence of both lighting variations (e.g., night and day) and moderate motion (such as that caused by driving).

Third, we are releasing the first publicly available driving dataset with simultaneous video and pulse oximeter recordings, both to allow for a fair comparison of future methods to our work and to enable further studies of how different sources of noise affect iPPG during driving.

While our proposed system achieves state-of-the-art performance in estimating vital signs during driving, the proposed system still struggles in presence of large motion. More improvements may be needed before the system can be deployed in a real driver monitoring system.

ACKNOWLEDGMENTS

Ashok Veeraraghavan and Ewa Nowara were partially supported by NSF Career Award IIS-1652633, NSF Expeditions Award CCF-1730574 and NIH Grant 5R01DK113269-03.

REFERENCES

[1] Ewa Magdalena Nowara, Tim K. Marks, Hassan Mansour, and Ashok Veeraraghavan. Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In *Computer Vision and Pattern Recognition (CVPR), 1st International Workshop on Computer Vision for Physiological Measurement*, 2018.

[2] Santokh Singh. Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Technical report, 2015.

[3] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE transactions on intelligent transportation systems*, 12(2):596–614, 2010.

[4] Deaths: Leading causes for 2014, author=Heron, Melonie, journal=National vital statistics reports, year=2016, publisher=Wiley Online Library volume=65, number=5,.

[5] Mitesh Patel, Sara KL Lal, Diarmuid Kavanagh, and Peter Rossiter. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert systems with Applications*, 38(6):7235–7242, 2011.

[6] KK Tripathi, CR Mukundan, T Lazar Mathew, et al. Attentional modulation of heart rate variability (hrv) during execution of pc based cognitive tasks. *Ind J Aerospace Med*, 47(1):1–10, 2003.

[7] Bing-Fei Wu, Yun-Wei Chu, Po-Wei Huang, Meng-Liang Chung, and Tzu-Min Lin. A motion robust remote-ppg approach to driver's health state monitoring. In *Asian Conference on Computer Vision*, pages 463–476. Springer, 2016.

[8] Jonny Kuo, Sjaan Koppel, Judith L Charlton, and Christina M Rudin-Brown. Evaluation of a video-based measure of driver heart rate. *Journal of safety research*, 54:55–e29, 2015.

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

[10] Jennifer A Healey and Rosalind W Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6(2):156–166, 2005.

[11] Paolo Napoletano and Stefano Rossi. Combining heart and breathing rate for car driver stress recognition. In *2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pages 1–5. IEEE, 2018.

[12] Shahina Begum. Intelligent driver monitoring systems based on physiological sensor signals: A review. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 282–289. IEEE, 2013.

[13] Lu Yu, Xianghong Sun, and Kan Zhang. Driving distraction analysis by ecg signals: an entropy analysis. In *International Conference on Internationalization, Design and Global Development*, pages 258–264. Springer, 2011.

[14] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

[15] Ming-Zher Poh, Daniel McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010.

[16] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.

[17] Ewa Magdalena Nowara, Ashutosh Sabharwal, and Ashok Veeraraghavan. Ppgsecure: biometric presentation attack detection using photoplethysmograms. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 56–62. IEEE, 2017.

[18] Lex Fridman, Philipp Langhans, Joonbum Lee, and Bryan Reimer. Driver gaze region estimation without use of eye movement. *IEEE Intelligent Systems*, 31(3):49–56, 2016.

[19] Ralph Oyini Mbouna, Seong G Kong, and Myung-Geun Chun. Visual analysis of eye state and head pose for driver alertness monitoring. *IEEE transactions on intelligent transportation systems*, 14(3):1462–1469, 2013.

[20] Taner Danisman, Ian Marius Bilasco, Chabane Djeraba, and Nacim Ihaddadene. Drowsy driver detection system using eye blink patterns. In *2010 International Conference on Machine and Web Intelligence*, pages 230–233. IEEE, 2010.

[21] Esra Vural, Mujdat Cetin, Aytul Ercil, Gwen Littlewort, Marian Bartlett, and Javier Movellan. Drowsy driver detection through facial movement analysis. In *International Workshop on Human-Computer Interaction*, pages 6–18. Springer, 2007.

[22] Paul Smith, Mubarak Shah, and Niels da Vitoria Lobo. Determining driver visual attention with one camera. *IEEE transactions on intelligent transportation systems*, 4(4):205–218, 2003.

[23] David Sandberg, Torbjörn Akerstedt, Anna Anund, Göran Kecklund, and Mattias Wahde. Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):97–108, 2010.

[24] Jongmin Yu, Sangwoo Park, Sangwook Lee, and Moongu Jeon. Driver drowsiness detection using condition-adaptive representation learning framework. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[25] Wikimedia Commons. File:solar spectrum en.svg — wikimedia commons, the free media repository, 2019. [Online; accessed 29-January-2020].

[26] Wikimedia Commons. File:low-pressure sodium lamp spectrum.svg — wikimedia commons, the free media repository, 2018. [Online; accessed 29-January-2020].

[27] Wikimedia Commons. File:xenon arc lamp profile.png — wikimedia commons, the free media repository, 2019. [Online; accessed 29-January-2020].

[28] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. Distanceppg: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express*, 6(5):1565–1588, 2015.

[29] Ewa Nowara and Daniel McDuff. Combating the impact of video compression on non-contact vital sign measurement using supervised learning. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[30] Bing-Fei Wu, Yun-Wei Chu, Po-Wei Huang, and Meng-Liang Chung. Neural network based luminance variation resistant remote-photoplethysmography for driver's heart rate monitoring. *IEEE Access*, 7:57210–57225, 2019.

[31] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[32] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[33] Mark van Gastel, Sander Stuijk, and Gerard de Haan. Motion robust remote-ppg in infrared. *IEEE Transactions on Biomedical Engineering*, 62(5):1425–1433, 2015.

[34] Ethan B Blackford and Justin 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.

[35] Weixuan Chen, Javier Hernandez, and Rosalind W Picard. Estimating carotid pulse and breathing rate from near-infrared video of the neck. *Physiological measurement*, 39(10):10NT01, 2018.

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

[37] Luis F Corral Martinez, Gonzalo Paez, and Marija 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.

[38] Peter Morante. Mesopic street lighting demonstration and evaluation final report. 2008.

[39] Byeonghoon Park, Yongchan Keh, Donghi Lee, Yongkwan Kim, Sungsoon Kim, Kisuk Sung, Jungkee Lee, Donghoon Jang, and Youngkwon 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.

[40] Claus Jaeger. Eye safety of ireds used in lamp applications. *Application note, OSRAM Opto Semiconductors GmbH*, 2009.

[41] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–10. IEEE, 2016.

[42] Sergey Tulyakov, Xavier Alameda-Pineda, Elisa Ricci, Lijun Yin, Jeffrey F Cohn, and Nicu 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.

[43] Ewout Van den Berg and Michael P Friedlander. Sparse optimization with least-squares constraints. *SIAM Journal on Optimization*, 21(4):1201–1229, 2011.

[44] Gareth H Loudon and Gina M Deininger. The physiological response during divergent thinking. *Journal of Behavioral and Brain Science*, 6(1):28–37, 2016.

Ewa M. Nowara received her B.S. degree in Physics from St. Mary's University in San Antonio, TX in 2015. She received the M.S. degree in Electrical and Computer Engineering from Rice University in 2018. Ms. Nowara is currently pursuing the Ph.D. degree in the Electrical and Computer Engineering Department at Rice University. She is interested in problems in computational imaging, computer vision, and machine learning with applications to imaging photoplethysmography.

Tim K. Marks (Member, IEEE) received the AB degree in Physics from Harvard University in 1991, the MS degree in Cognitive Science from the University of California, San Diego in 2001, and the PhD degree in Cognitive Science from the University of California, San Diego (UCSD) in 2006. Dr. Marks is a Senior Principal Research Scientist at the Mitsubishi Electric Research Laboratories (MERL), in Cambridge, MA, where he has worked since 2008. From 2006–2008, he was a postdoctoral researcher in robotics at the UCSD Department of Computer Science, in collaboration with the computer vision group at the NASA/Caltech Jet Propulsion Laboratory. Before attending graduate school, he was a high school math and physics teacher and then a mathematics textbook editor for Houghton Mifflin. His current research interests include computer vision and machine learning. He has published more than 50 papers in journals and conferences (including the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, CVPR, ICCV, ECCV, IROS, and NeurIPS), as well as a college Number Theory textbook. He has been granted 16 US patents.

Hassan Mansour (S'99, M'09, SM'17) received the B.E. degree in computer and communications engineering from the American University of Beirut, Beirut, Lebanon, in 2003, and the M.A.Sc. degree in electrical and computer engineering and the Ph.D. degree in electrical and computer engineering from The University of British Columbia, Vancouver, BC, Canada, in 2005 and 2009, respectively. Between January 2010 and January 2013, he was a Postdoctoral Research Fellow with the Department of Computer Science, the Mathematics Department, and the Department of Earth, Ocean, and Atmospheric Sciences, The University of British Columbia. He is currently a Senior Principal Research Scientist with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. His research interests are in inverse problems, compressed sensing, sparse signal reconstruction, image enhancement, and scalable video compression and transmission. His current research is focused on the design of efficient acquisition schemes and reconstruction algorithms for natural images, radar sensing, video analytics, and inverse scattering problems. Dr. Mansour is a member of the IEEE Computational Imaging Technical Committee and the IEEE Sensor Array and Multichannel Technical Committee. He is also an Associate Editor for the IEEE Transactions on Signal Processing.

Ashok Veeraraghavan received the bachelor's degree in electrical engineering from the Indian Institute of Technology, Madras, Chennai, India, in 2002 and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA, in 2004 and 2008, respectively. He is currently a Professor of Electrical and Computer Engineering at Rice University, Houston, TX, USA. Before joining Rice University, he spent three years as a Research Scientist at Mitsubishi Electric Research Labs, Cambridge, MA, USA. His research interests are broadly in the areas of computational imaging, computer vision, machine learning, and robotics. Dr. Veeraraghavan's thesis received the Doctoral Dissertation Award from the Department of Electrical and Computer Engineering at the University of Maryland. He was the recipient of the National Science Foundation CAREER Award in 2017. At Rice University, he directs the Computational Imaging and Vision Lab.

Near-Infrared Imaging Photoplethysmography During Driving: Supplementary Material

Ewa M. 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

I. Eye safety calculations

To ensure that our NIR system is within the eye safety standards, we conducted careful computations, according to the OSRAM eye safety note [6].

A. Cornea safety

Cornea burn hazard is concerned with the total amount of light per unit area reaching the eye at a given distance from the light source. We conducted the computations at a distance of 3 inches as the worst-case scenario distance from the lights to the face of the subject in the car. The maximum safe irradiance for the cornea is 100 Wm^{-2} for an area of 1 m^2. We convert this limit by taking into account the surface area of the light meter sensor. The diameter of the sensor was 9.5 mm and the area of the sensor is approximately $(0.5 * 9.5 \text{ mm})^2 * \pi$ = $7{*}10{-}5$ m^2. Therefore, the maximum safe total irradiance that can reach the pupil is 7 mW. We measured the total irradiance of all the LEDs together at a distance of 3 inches from the light meter sensor to be 2.38 mW, which is significantly below the safety limit of 7 mW.

B. Retina safety

The retinal thermal hazard is additionally concerned with how narrow the beam of the light is. For example, narrow-beam lasers are more dangerous for the retina than broad-beam LEDs.

We measure the maximum radiance that is safe for the retina using the following equation:

$$L_{IR} = \frac{I_e * R(\lambda)}{(\frac{l+w}{2})^2 * \#\text{ LEDs}}\,. \quad (1)$$

Here, I_e is the radiant intensity per solid angle in mW sr^{-1}. $R(\lambda)$ is a wavelength-dependent burn hazard weighing function. For the 900–950 nm wavelength range it is 0.3–0.4, so we use $R(\lambda)$ = 0.4 as the worst-case scenario. Here, l and w are the length and width of the light source in mm, respectively. Each of the LEDs used in our illuminators was 1 mm $\times$ 1 mm. The maximum radiant intensity is computed for an individual LED, therefore, we divide the total measured radiant intensity by the total number of the LEDs used (4 in each illuminator).

To compute the solid angle, given the area of our light meter sensor (A) and the distance between the light source and the light meter (d), we use the following equation:

$$\Omega = \frac{A}{d^2} = \frac{(7 * 10^{-5}\ \mathrm{m}^2)}{0.076\ \mathrm{m}}$$

The radiant intensity I_e is $\frac{2.38\ \mathrm{mW}}{0.012\ \mathrm{sr}} = 198.33\ \mathrm{mW\ sr}^{-1}$, where 2.38 mW is the total power of the LEDs. Therefore, the maximum radiant intensity becomes:

$$L_{IR} = \frac{198.33\mathrm{mW\ sr}^{-1} * 0.4}{(\frac{1+1}{2})^2 * 4} = 19.83\ \mathrm{mW\ mm}^{-2}\ \mathrm{sr}^{-2} \quad (2)$$

which is significantly below the recommended retina safety limit of 545.5 mW mm^{-2} sr^{-2}.

II. Compared Algorithms

We now provide implementation details of the state-of-the-art algorithms that we evaluated on our MR-NIRP Car Dataset to compare against our proposed AutoSparsePPG method.

1) **SparsePPG** [1] - algorithm using robust principal components analysis (RPCA), followed by sparse spectrum estimation similar to AutoSparsePPG, but using fixed optimization parameters set a priori. We performed leave-one-subject-out cross validation to find the best optimization parameters for SparsePPG on MR-NIRP Indoor dataset: $\gamma = 0.5$ and $r = 25$ rank for RPCA if 48 mean regions were used or $r = 3$ when 5 median regions were used, $\lambda = 0.1$ and $\mu = 0.2$ for sparse spectrum estimation, $\alpha = 0.8$ for time window fusion. We used the same parameters for the MR-NIRP Car dataset because it was very expensive to search over the parameter space (one of the main shortcomings of the SparsePPG method, which is no longer a problem with our proposed AutoSparsePPG method).
2) **DistancePPG** [2] - algorithm using a weighted average of different facial regions, where the weights are set based on the SNR of the signals computed as the ratio of the area under the curve of the power spectrum around the maximum frequency peak to the area under the rest of the spectrum. Regions that have SNR below 0.2, or absolute amplitude above a fixed threshold, or the difference in estimated heart rate between the current and the previous time window larger than 12 bpm are removed from each time window before averaging. We used the amplitude threshold set as four times the mean amplitude of normalized iPPG signals for all videos in each dataset, computed separately in NIR and RGB.

TABLE I
HR ESTIMATION ERRORS ON THE MR-NIRP CAR DATASET (ADDITIONAL HEAD MOTION)

	Driving Day		**Driving Night**		**Driving All**		**Garage**	
	NIR	**RGB**	**NIR**	**RGB**	**NIR**	**RGB**	**NIR**	**RGB**
	PTE6 [%] (higher is better)							
AutoSparsePPG	**24.5 ± 4.2**	26.1 ± 4.0	**16.7 ± 7.0**	10.4 ± 2.7	**22.9 ± 3.6**	24.2 ± 3.8	21.6 ± 2.9	50.0 ± 4.7
SparsePPG [1]	23.6 ± 5.5	26.6 ± 5.9	8.7 ± 6.3	5.1 ± 2.5	20.5 ± 4.7	24.1 ± 5.5	**26.4 ± 4.6**	30.0 ± 6.9
DistancePPG [2]	**24.5 ± 3.9**	17.5 ± 3.8	14.9 ± 3.3	9.0 ± 1.6	22.5 ± 3.2	16.5 ± 3.4	24.1 ± 2.9	41.8 ± 5.6
ICA [3]	N/A	45.7 ± 5.0	N/A	12.4 ± 7.5	N/A	41.8 ± 5.2	N/A	60.8 ± 5.7
CHROM [4]	N/A	**50.6 ± 3.8**	N/A	**29.7 ± 11.9**	N/A	**48.1 ± 3.8**	N/A	**64.1 ± 5.6**
POS [5]	N/A	22.2 ± 2.3	N/A	12.0 ± 1.6	N/A	21.0 ± 2.2	N/A	26.7 ± 3.2
	RMSE [bpm] (lower is better)							
AutoSparsePPG	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	14.8 ± 1.5
SparsePPG [1]	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM
DistancePPG [2]	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM	14.4 ± 2.2
ICA [3]	N/A	**12.8 ± 1.7**	N/A	> 15 BPM	N/A	13.7 ± 1.7	N/A	**9.8 ± 1.8**
CHROM [4]	N/A	13.0 ± 1.2	N/A	> 15 BPM	N/A	**13.6 ± 1.2**	N/A	10.0 ± 1.9
POS [5]	N/A	> 15 BPM	N/A	> 15 BPM	N/A	> 15 BPM	N/A	> 15 BPM

TABLE II
HR ESTIMATION ERRORS ON THE MR-NIRP CAR DATASET (MINIMAL HEAD MOTION) WITH AUTOSPARSEPPG VARIATIONS

	PTE6 [%] (higher is better)				**RMSE [bpm] (lower is better)**			
	Driving		Garage		Driving		Garage	
	NIR	RGB	NIR	RGB	NIR	RGB	NIR	RGB
AutoSparsePPG median ROIs + OP	61.0 ± 5.2	**31.5 ± 3.1**	**81.9 ± 5.9**	**91.1 ± 1.9**	**11.6 ± 1.8**	> 15 BPM	**5.1 ± 1.4**	**2.9 ± 0.4**
AutoSparsePPG median ROIs + OP + RPCA	**61.5 ± 5.0**	30.9 ± 2.9	80.7 ± 5.8	89.9 ± 2.2	11.7 ± 1.7	> 15 BPM	5.8 ± 1.4	3.1 ± 0.5
AutoSparsePPG median ROIs + RPCA	32.9 ± 3.5	27.7 ± 3.4	70.1 ± 5.6	90.3 ± 2.8	> 15 BPM	> 15 BPM	7.3 ± 1.38	3.3 ± 0.7
SparsePPG median ROIs + RPCA [1]	11.5 ± 2.9	15.8 ± 2.54	14.4 ± 5.4	70.8 ± 7.3	> 15 BPM	> 15 BPM	> 15 BPM	> 15 BPM
AutoSparsePPG mean ROIs + OP + RPCA	53.9 ± 4.4	26.4 ± 2.9	72.7 ± 5.7	81.0 ± 5.2	> 15 BPM	> 15 BPM	8.2 ± 1.6	7.0 ± 2.2
SparsePPG mean ROIs + RPCA [1]	17.4 ± 3.4	21.5 ± 3.2	35.6 ± 6.8	53.6 ± 9.8	> 15 BPM	> 15	> 15 BPM	> 15 BPM

3) **ICA** [3] - independent component analysis applied to the spatially averaged R, G, B channels, followed by detrending of the signals.
4) **CHROM** [4] - chrominance features computed from the spatially averaged R, G, B channels.
5) **POS** [5] - projection of iPPG signals computed from the spatially averaged R, G, B channels onto the plane in color space that is orthogonal to the skin tone.

The open-source implementations of [3], [4], and [5] can be found in [7].

III. Additional Results During Driving

There were often large and sudden movements of the head caused by the motion of the car, even when the participant was trying to sit still (the *minimal head motion* condition). This large motion sometimes makes it very challenging to recover the iPPG signal, especially in single-channel NIR. When this involuntary head motion due to driving was combined with head motion caused by the subject looking around and talking (the *additional head motion* condition), the total amount of head motion was quite large. Consequently, all methods performed comparably poorly with *additional head motion*, even in RGB, with PTE6 $<$ 30% for most methods. See Table I.

IV. AutoSparsePPG variations

We compared the performance of SparsePPG [1] and AutoSparsePPG under different pre-processing conditions and different denoising methods in Table II. We compared the performance using iPPG matrices computed with 48 mean facial regions as used in [1] and five median regions to test how much robustness is gained by using the median regions. Using the five larger median regions consistently performed better than using the 48 smaller mean regions. In order to evaluate how much improvement was offered by denoising with orthogonal projections (OP), we also compared the performance of our proposed method AutoSparsePPG with OP to AutoSparsePPG with RPCA instead of OP, and AutoSparsePPG with using RPCA and OP jointly, each followed by sparse spectrum estimation. We found that further denoising the iPPG signals with RPCA before applying sparse spectrum estimation often completely removed the iPPG signal when OP was also used. To make sure that most of the signal was not removed with RPCA, we only applied RPCA when the signal norm was above a fixed threshold of 10^{-3}.

We found that using 5 median regions consistently performed better than using 48 mean regions in the presence and the absence of motion. Moreover, we found that AutoSparsePPG with OP and RPCA used jointly performs similarly to AutoSparsePPG with OP without RPCA. Therefore, we conclude that applying RPCA is not worth the additional computational complexity, and high accuracy can be achieved with only OP.

V. Hardware Variations

We now present two hardware additions that we tested to further reduce the ambient light variations during driving. First, we compared using bandpass filters with different passband wavelengths to account for blueshift effects. Second, we used “dark frame subtraction” to further reduce ambient light

TABLE III
HR ESTIMATION ERRORS ON THE MR-NIRP CAR DATASET (MINIMAL HEAD MOTION) WITH AUTOSPARSEPPG IN DIFFERENT HARDWARE SETTINGS.

	Driving		Garage	
	PTE6 (%)	RMSE (bpm)	PTE6 (%)	RMSE (bpm)
940 $\pm$ 5 nm bright	**61.0 $\pm$ 5.2**	**11.6 $\pm$ 1.8**	**81.9 $\pm$ 5.9**	**5.1 $\pm$ 1.4**
940 $\pm$ 5 nm subtracted	38.6 $\pm$ 5.0	$>$ 15 BPM	78.4 $\pm$ 5.3	7.2 $\pm$ 2.2
975 $\pm$ 25 nm bright	58.2 $\pm$ 5.3	$>$ 15 BPM	76.0 $\pm$ 6.0	8.6 $\pm$ 2.3
975 $\pm$ 25 nm subtracted	47.3 $\pm$ 5.6	$>$ 15 BPM	73.7 $\pm$ 6.4	9.7 $\pm$ 3.2

variations. However, these innovations did not yield significant improvements in our performance over the original hardware setup presented in [1].

A. *Blueshift in optical filters*

Optical filters, especially interference filters, are usually designed for small angles of incidence. When the angle of incidence of the incoming light is increased from zero, there is a reduction in the path length difference between the transmitted and the reflected beams, making the destructive interference of the light waves incident on the filter less effective. This results in a shift towards shorter transmitted wavelengths, referred to as *blueshift*. The larger the angle of incidence, the larger the blueshift effect: $\lambda = \lambda_0\sqrt{1 - (sin(\frac{\theta}{n}^2))}$, where λ is the transmitted wavelength, λ_0 is the incident light wavelength, θ is the angle of incidence, and n is the refractive index of the filter [8]–[11].

In a car environment, it is difficult to control the angle of incidence of the light reflecting off the face and reaching the camera. We expect that the range of angles of incidence with our car setup will be between 0–30 degrees. This could potentially lead to blueshift and allow wavelengths shorter than 930 nm to be transmitted, making the 940 nm bandpass filter inefficient at blocking the ambient light. We analyzed how much the blueshift might affect our measurements by recording the sunlight spectrum with a bandpass filter placed in front of a spectrometer (Ocean Optics USB2000+) and gradually increasing the angle of the filter with respect to the spectrometer sensor. As shown in Fig. 1, changing the angle of the filter by 20 degrees or more shifted the transmitted wavelength by as much as 30 nm. Therefore, instead of transmitting light within the expected narrow range of 935–945 nm as expected, we are instead transmitting light closer to 910–920 nm, which is no longer ideal for reducing the sunlight effects. To account for these blueshift effects and to ensure we can block the light in the <930 nm range, we repeated the spectrometer measurements and video recordings of all subjects with a 975 nm bandpass filter. We used a wider bandwidth of 50 nm on the 975 nm filter to allow more light, to improve the iPPG SNR which decreases with increasing wavelength in NIR. The 975 nm filter could block the ambient sunlight better at larger angles of incidence. However, we found that the 975 ± 25 nm filter did not provide improved performance over the 940 ± 5 nm filter, as shown in Table III, suggesting that if blueshift was present, its effects on the quality of iPPG signals in the presence of varying sunlight were minimal.

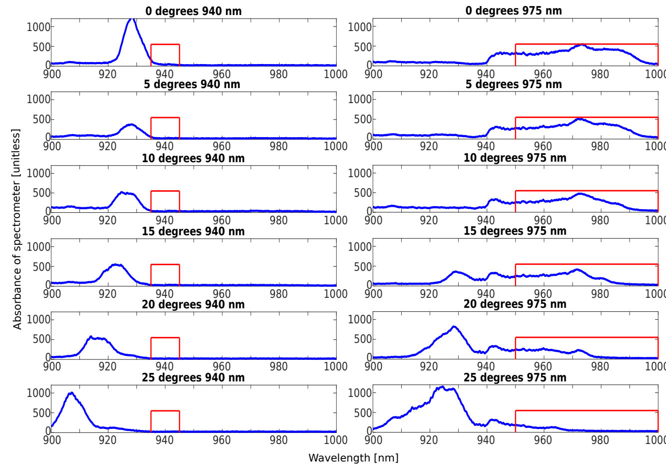


Fig. 1. Spectrometer measurements of blueshift effects with 940 nm and 975 nm filters with varying angles of incidence.

B. *Dark frame subtraction*

Even with a narrow bandpass filter on the camera, there may be remaining outside light variations on the face when the sunlight is very strong. To further reduce these light variations, we used an approach referred to as ”dark frame subtraction,” often used for reducing noise in astronomy imaging [12]. The camera frame capture is synchronized with the light source, such that only every other frame is illuminated by the LEDs (“bright frames”). The remaining camera frames are captured when the LEDs are turned off (“dark frames”) and are only illuminated by the ambient light. To reduce the ambient light, the dark frames are subtracted from the adjacent bright frames. We recorded NIR images at 60 frames per second (fps), however, if we use only the dark-frame-subtracted frames or only the bright frames, the recordings are effectively at 30 fps. An example of reducing ambient light with dark frame subtraction is illustrated in Fig. 2. Notice that in the subtracted frame, the face mostly illuminated by the LEDs is visible but the background region illuminated by the ambient light is removed. Despite reducing the ambient light variations, we found that dark frame subtraction did not improve the HR estimation results, as shown in Table III. In the table, “bright” refers to using only the bright frames (equivalent to using constant NIR illumination while recording at 30 fps), while “subtracted” refers to using the 30 fps video obtained by subtracting each dark frame from the adjacent bright frame.

From the results, it is clear that rather than improving the algorithm’s performance, using dark frame subtraction did not help, and it actually reduced the performance significantly in the driving condition. We now discuss why we believe dark frame subtraction did not help. Dark frame subtraction is predicated on the assumption that the ambient light is approximately constant in the 1/60 second between each bright frame and the adjacent dark frame. When driving outdoors, this assumption does not hold, because the ambient lighting (e.g., lighting on the face when driving through the shadow

Fig. 2. Illustration of dark frame subtraction: “bright” frames illuminated with the NIR LEDs are subtracted from “dark frames” captured when the NIR LEDs were turned off. A large portion of the ambient light is reduced with dark frame subtraction, as shown inside the red region.

of tree branches or driving past a streetlamp) often changes quite rapidly.

References

[1] Ewa Magdalena Nowara, Tim K. Marks, Hassan Mansour, and Ashok Veeraraghavan. Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In *Computer Vision and Pattern Recognition (CVPR), 1st International Workshop on Computer Vision for Physiological Measurement*, 2018.

[2] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. Distanceppg: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express*, 6(5):1565–1588, 2015.

[3] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

[4] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[5] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[6] Claus Jaeger. Eye safety of ireds used in lamp applications. *Application note, OSRAM Opto Semiconductors GmbH*, 2009.

[7] Daniel McDuff and Ethan Blackford. iphys: An open non-contact imaging-based physiological measurement toolbox. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6521–6524. IEEE, 2019.

[8] Laurent Frey, Lilian Masarotto, Marilyn Armand, Marie-Lyne Charles, and Olivier Lartigue. Multispectral interference filter arrays with compensation of angular dependence or extended spectral range. *Optics express*, 23(9):11799–11812, 2015.

[9] PH Lissberger and WL Wilcock. Properties of all-dielectric interference filters. ii. filters in parallel beams of light incident obliquely and in convergent beams. *JOSA*, 49(2):126–130, 1959.

[10] Thomas D Rahmlow, Markus Fredell, Sheetal Chanda, and Robert Johnson. Ultra-narrow bandpass filters for infrared applications with improved angle of incidence performance. In *Advanced Optics for Defense Applications: UV through LWIR*, volume 9822, page 982211. International Society for Optics and Photonics, 2016.

[11] H Angus Macleod and H Angus Macleod. *Thin-film optical filters*. CRC press, 2010.

[12] Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Entropy-based dark frame subtraction. In *PICS*, pages 293–298, 2001.