ABSTRACT

Title of Thesis: Compression and Multi-Spectral Sensing
for Video Based Physiological Monitoring

Carl Steinhauser
Master of Science, 2022

Thesis Directed by: Dr. Min Wu
Department of Electrical and Computer Engineering

Remote physiological monitoring is an active area of research that extends monitoring capabilities traditionally found in a clinical setting towards the home, telehealth, and beyond. In particular, there is interest in leveraging consumer electronic devices for sensing physiological characteristics such as heart rate, heart rate variability, and blood oxygen saturation. This thesis focuses on enhancing the understanding and usage of the sensing component for these applications to improve the performance and quality of cardio-physiological monitoring. First, a close relationship between the color spaces used for video compression and the color projection planes commonly used for heart rate estimation is identified. The study demonstrates the impact of this observation on real and synthetic data to provide a foundation to guide future video coding to optimize its configurations to better preserve the heart rate signal for health related applications. Second, an investigation with a commercial-off-the-shelf (COTS) multi-spectral sensor is presented with key observations related to the sampling rate, exposure settings, and multi-channel processing. These observations will enable better usage of the sensor for future studies and data collections that leverage the more precise spectral measurements from the multi-spectral sensor compared to standard RGB cameras.

COMPRESSION AND MULTI-SPECTRAL SENSING FOR VIDEO BASED PHYSIOLOGICAL MONITORING

by

Carl Frederick Steinhauser

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2022

Advisory Committee:
Professor Min Wu, Chair/Advisor
Professor Yang Tao
Assistant Professor Chai-Wai Wong

Acknowledgments

I would like to express my gratitude towards my advisor, Professor Min Wu, who has helped guide my Master's Thesis work. Her critical and constructive feedback guided the research and honed my abilities to communicate the technical results and impact of the work.

I would like to thank my Master's committee members, Professor Yang Tao and Professor Chau-Wei Wong. Their review and feedback are greatly appreciated.

I would also like to thank my fellow students in the MAST research group for both their comments and feedback on my presentations in group meetings and for giving me the opportunity to learn about their research.

Lastly, I want to thank Stephanie and the rest of my family for their support during the completion of my graduate degree. The time and effort of classes and research in addition to work have created a busy schedule during this season of life. My success wouldn't be possible without their support and encouragement.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AVC   Advanced Video Coding
COTS   Commercial of the Shelf
CRF   Constant Rate Factor
DFT   Discrete Fourier Transform
HEVC   High Efficiency Video Coding
HR   Heart Rate
HVS   Human Visual System
NIR   Near Infrared
POS   Plane-orthogonal to the skin
PPG   Photoplethysmography
RoR   Ratio-of-Ratios
rPPG   Remote Photoplethysmography
SpO_2 Peripheral Blood Oxygen Saturation
SNR   Signal-to-Noise Ratio

Chapter 1

Introduction

1.1 Overview

Remote photoplethysmography (rPPG) is an active area of research that began with heart rate (HR) estimation in a controlled setting and has been extended to less controlled settings as well as estimation of other parameters such as HR variability and blood oxygen saturation (SpO_2). Within this area of research, there is a focus on demonstrating the capabilities of commercial sensors, such as cameras on consumer phones. Physiological monitoring with commercial devices that are common in most homes creates the opportunity for people to more easily monitor vitals while away from a standard clinical setting. The following expands this body of work with a focus to enhance the understanding and usage of sensing to improve the performance and quality of cardio-physiological monitoring.

The outline of this thesis is as follows. In this chapter, relevant research that provides foundations of video-based physiological monitoring will be reviewed as it relates to the physiological source of the signal, the sensor, and algorithms to estimate the physiological state (Figure 1.1). Chapters 2 and 3, contain the details of the main contributions of this thesis work that focuses on the sensor component of the process. Chapter 2 details the original finding of this work that

1

Figure 1.1: High level outline of remote physiological monitoring with emphasis on the sensor, the focus of the thesis work.

identified the close relationship between the color space used in standard video compression and the physiological signal extracted by plane orthogonal to the skin (POS), a first principles motivated rPPG algorithm. This observation is further explored and the concept of adjusting video compression settings when the application is known to be rPPG is introduced and experimentally demonstrated. Chapter 3 describes the building blocks developed under this thesis work for using the Sentera 6x sensor, a commercial off-the-shelf (COTS) multi-spectral sensor, for future research related to physiological monitoring. In this section, demonstrations with the Sentera sensor detail the expected behavior of rPPG based on a 5 Hz frame rate, a demonstration of multi-spectral frame alignment to streamline the data processing pipeline, and a characterization of the automatic exposure controller in the Sentera 6x Sensor. Finally, Chapter 4 summarizes the work and discusses future research opportunities enabled by the contributions in this thesis.

1.2   Background on rPPG

1.2.1   Physiological Signal Source

Video-based physiological monitoring is primarily based on detecting changes in the color of the skin surface caused by blood volume variation as the heart beats. In traditional photoplethysmography (PPG) that is often evaluated at the fingertip, the transmission of a light source through the skin is used to measure this blood volume variation which has the temporal behavior as shown
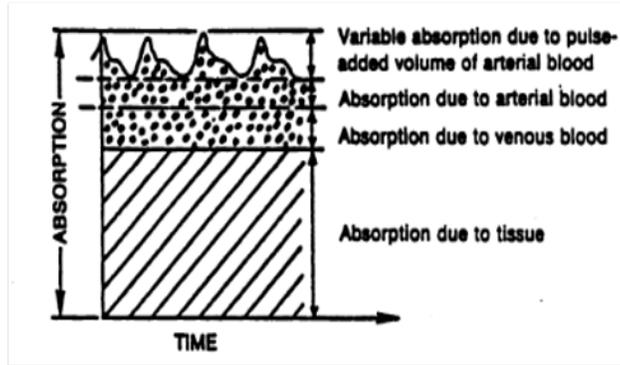
Figure 1.2: Visualization of the PPG signal and sources of signal attenuation including tissue, venous blood, and arterial blood (from [1]).
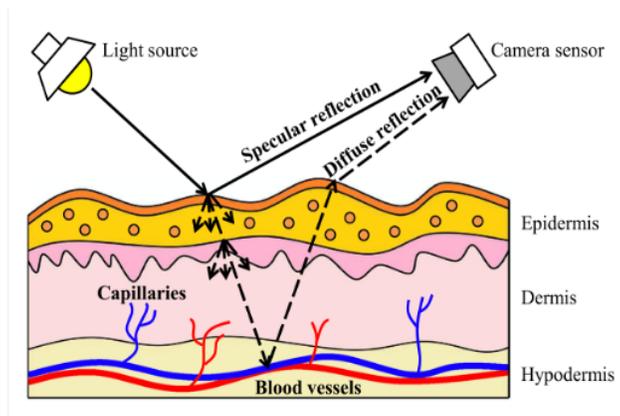
Figure 1.3: Optical path of light for rPPG scenario where blood volume variation impacts the diffuse reflection observed at the camera sensor (from [2]).

in Figure 1.2. While the physiological source of the HR signal is the same with video-based measurement, the path of the light carrying the signal is different. In this case, the measurement is reflective where the ambient (or manual light fixture) transmits to the skin, penetrates to the blood vessels, reflects back through the skin, and transmits to the sensor aperture (Figure 1.3). While the light path is different for reflective PPG, the same principles apply to the video-based sensing methodologies with consumer electronic devices that have been more recently researched.

Measuring the pulse in a single spectral bandwidth is enough to estimate the HR or HR variability. If two appropriate bandwidths are measured, the blood oxygen saturation (SpO$_2$) can also
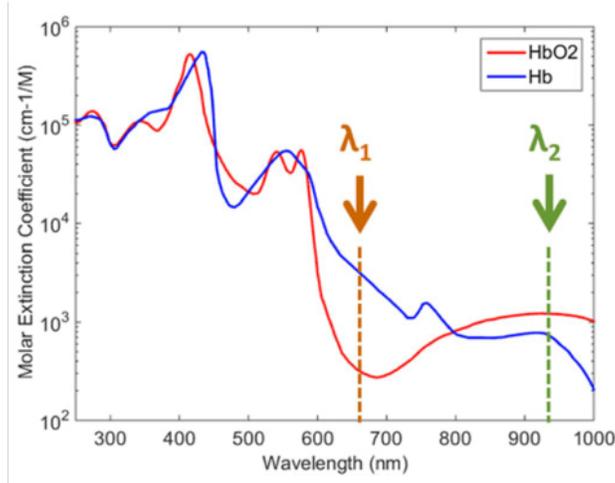
Figure 1.4: Extinction coefficient of oxygenated (HbO$_2$) and deoxygenated (Hb) hemoglobin as a function of wavelength with indicators at common wavelengths for traditional pulse oximeters (660 and 940 nm) (from [3]).

be estimated. The physical property that enables this measurement is the variation of extinction coefficients as a function of wavelength for oxygenated and de-oxygenated hemoglobin in the blood (Figure 1.4).

The most common method for estimating SpO$_2$ with traditional pulse oximeters is the ratio-of-ratio (RoR) method. This method uses the AC and DC components of the HR signal at two wavelengths to isolate information in the signal related to the oxygenated and de-oxygenated hemoglobin, HbO$_2$ and Hb, respectively. This approach cancels out unknown terms that are unrelated to the blood oxygen saturation such as absorption from skin tissue and venous blood.

$$RoR = \frac{AC_{\lambda_1}/DC_{\lambda_1}}{AC_{\lambda_2}/DC_{\lambda_2}} \tag{1.1}$$

A linear function of RoR is often fitted to estimate SpO$_2$. The exact derivation does not result in a precisely linear function, but a linear fit is an effective approximation. Common wavelengths used in commercial fingertip pulse oximeters are 660 nm and 940 nm with narrow band detectors.

4
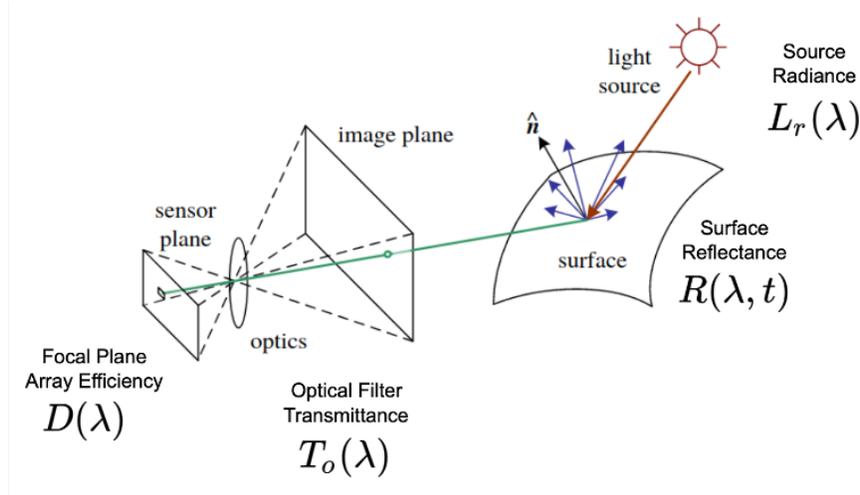
Figure 1.5: Optical path of reflective rPPG signal with consideration of wavelength (adapted from [5]).

Further details on the derivation of the RoR method are available in "Design of Pulse Oximeters" by J.G. Webster [4].

1.2.2 Sensor

Before discussing aspects of the sensor relevant to rPPG, we will expand upon the prior figures and consider behavior more precisely as a function of wavelength as shown in Figure 1.5. In this figure, the skin reflectance that is the source of the rPPG signal is simplified to be a time-varying reflectance, $R(\lambda, t)$. In this schematic, the sensor model is composed of the focal plane array efficiency, $D(\lambda)$, and the optical filter transmittance, $T_o(\lambda)$. The radiance of the light source is $L_r(\lambda)$. This results in the observed signal at a particular pixel, $x_{i,j}(t)$, being proportion to

$$x_{i,j}(t) \propto \int L_r(\lambda) R(\lambda, t) T_o(\lambda) D(\lambda) d\lambda \qquad (1.2)$$

RGB cameras are primarily used to capture images that are intended to be displayed and to

5
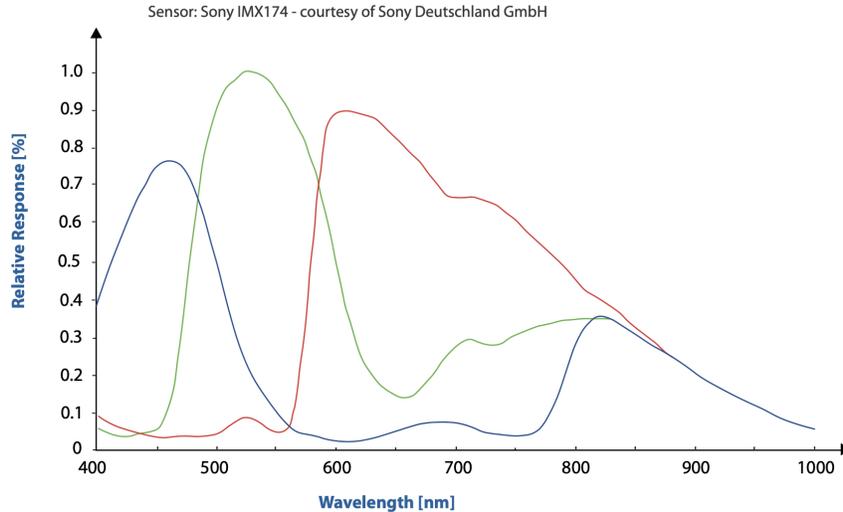
Figure 1.6: Responsivity of color channels for Sony IMX174 Detector (from [6]). The figure should be viewed in color.

look like the original observation. To achieve this, the cameras are configured to have relatively broad spectral responsivity across the visible spectrum (400-700nm). Most consumer cameras achieve multi-color observations with a Bayer filter which impacts $D(\lambda)$ for pixels that are then de-mosaiced to produce the RGB image displayed on a screen or printed to a paper. A representative example of the RGB spectral responsivity of a Bayer filter is shown in Figure 1.6 where the bandwidth of each color is on the order of 100 nm.

In contrast to consumer cameras, it is common in experimental data collections to have a panchromatic camera configured with an optical band-pass filter aligned in the optical path. If this experimental configuration has multiple colors, it is likely the detector response, $D(\lambda)$ does not vary, but the optical filter transmittance, $T_o(\lambda)$ does vary with each color channel [7, 8]. Optical filters can have pass-bands on the order of 10 nanometers or smaller.

While the source of the spectral filtering within the camera has minimal effect on the observed rPPG signal, the width of the band-pass filter can be important particularly if the application extends beyond HR estimation and is moving towards SpO_2 estimation. The primary benefit of a

narrow spectral response is reduced dependence on the source radiance, $L_r(\lambda)$, which can vary significantly between different natural and artificial light sources.

The last piece of the sensor that is important to this work is data storage. Scientific sensors often store data in an uncompressed or lossless compression format to preserve all of the information. Consumer video cameras usually do not have either the data storage capacity or the data throughput capacity to store uncompressed videos, so the only available output is compressed video. This attribute will be a core focus of the work presented in Chapter 2.

1.2.3 rPPG Algorithm Processing

Heart rate estimation with consumer devices was first demonstrated with a contact-based methodology used in the experimental setup with the subject's finger placed on the camera [9, 10]. Later work extended to remote photoplethysmography (rPPG) by estimating heart rate from videos of the subject's face [11, 12]. These experiments relied on the reflective signal source visualized in Figure 1.3. Multiple algorithmic approaches have been applied to this problem such as blind source separation (BSS) [11, 12], chrominance based rPPG (CHROM) [13], the plane orthogonal to the skin (POS) method [2], and data-driven approaches [14, 15]. Many of these methods leverage information in all of the red, green, and blue color channels and focus on changes in color with the intent to limit the influence of changes in luminance that are more likely to be caused by adjustment of ambient light or movement.

A specific algorithm used in this thesis work is the POS method [2] which was selected due to the derivation being motivated by first principles. This algorithm starts with an RGB time series, $\mathbf{c}(t)$, that has been extracted from a sequence of video frames. The algorithm uses the projection

basis

$$P_{POS} = \begin{bmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix} \tag{1.3}$$

to project the RGB signal with the intent of removing the intensity component.

$$\mathbf{s}(t) = P_{POS}\,\mathbf{c}(t) \tag{1.4}$$

Next, a weighted average of the two remaining dimensions of the time series $\mathbf{s}(t)$ is computed as

$$\hat{p}_{POS}(t) = s_1(t) + \alpha\,s_2(t) \tag{1.5}$$

where $s_i(t)$ is the ith dimension of $\mathbf{s}(t)$, $\alpha = \sigma(s_1)/\sigma(s_2)$, and $\sigma(x)$ is the standard deviation [2].

The color space projection in the POS algorithm will be of significant importance in Chapter 2.

Chapter 2

Video Compression and Cardiac Monitoring

In this chapter, a close connection between video compression and the specific projection plane used in the POS algorithm, a first-principles motivated algorithm, is identified. This connection motivates the consideration of alternative color spaces to be used for video compression when the application is known to be physiological monitoring. Improved rPPG signal preservation at low bitrates was demonstrated on synthetic and real data using the RGB24 color space during compression compared to YUV444 and YUV420. These results indicate an opportunity to reduce errors in rPPG applications by considering the video compression color space as a tunable parameter in the system design.

2.1 Related Work

Prior work studying the relationship between video compression and rPPG has investigated the relationship between the constant rate factor (CRF) and the SNR and the error of the HR signal [16, 17]. The CRF configuration parameter impacts the resulting bit rate of the compressed video with low CRF factors having smaller compression rates. Additionally, Zhao et al. illustrated varying rates of performance degradation with different rPPG algorithms indicating some algorithms are more robust to the artifacts introduced by video compression [17]. Nowara et al.

evaluated the behavior of neural networks trained to extract the rPPG signal and the behavior at varying compression levels as well as the behavior when test data compression rates vary from the training data [18].

These prior works have evaluated video compression as it relates to HR estimation with the intent of understanding behavior at varying bit rates and other default compression configuration parameters. This chapter extends the scope to consider how video compression can be configured to better preserve the rPPG signal at constant bit rates. The direct amplitude error is considered instead of through a surrogate metric of SNR in order to inform studies beyond HR estimation such as SpO_2 estimation, where the amplitude entails the physiological signal.

2.2   Review of Color Spaces for Video Compression

There are many color spaces used for both storing images and videos as well as displaying them. Historically, the YUV color space that was standardized for cathode ray tube displays, originally only had the luminance (Y) color channel when the transmission was grayscale. Later, the chrominance (U and V) color channels were introduced to limit the additional bandwidth required to include color [19].

The more modern variant of YUV is YCbCr which is not as influenced by the cathode ray tube history and includes a DC shift to bring the values back into the range of [0, 255] that is used for 8-bit digital storage. However, the name "YUV" can still exist in some references and documentation that are referring to YCbCr. While the names are different, the principle of a single channel for luminance and two chrominance channels is the same. YCbCr is a parametric color

space with the basis vectors defined by

$$B_{YCbCr}^{T} = \begin{bmatrix} \mathbf{b}_Y^T \\ \mathbf{b}_{Cb}^T \\ \mathbf{b}_{Cr}^T \end{bmatrix} = \begin{bmatrix} K_R & K_G & K_B \\ -\frac{K_R}{2(1-K_B)} & -\frac{K_G}{2(1-K_B)} & \frac{1}{2} \\ \frac{1}{2} & -\frac{K_G}{2(1-K_R)} & -\frac{K_B}{2(1-K_R)} \end{bmatrix} \qquad (2.1)$$

where $K_R + K_G + K_B = 1$. Note that the Y channel contains a weighted sum of the RGB color channels while the Cb and Cr channels contain information about differences. Particular standards have varying definitions for the K coefficients [20], [21], [22]. It is also worthwhile to note this is not necessarily an orthogonal basis for all valid K coefficients. The YCbCr color space is commonly used in the compression of images and videos with formats such as JPEG, JPEG2000, Advanced Video Coding (AVC), and High Efficiency Video Coding (HEVC) [19].

When applied to image and video compression, the color space transformation from RGB to YCbCr is also often combined with chrominance sub-sampling of the Cb and Cr color channels. The human visual system (HVS) is not as sensitive to the information in the chrominance channels compared to the luminance channel, and this property enables the downsampling to significantly decrease necessary bandwidth with limited perceptual difference [19]. This sub-sampling is an important property for the following analysis and it is important to note that the design decision of a conventional video coding is motivated by the human visual system, not to preserve such inherent information of a physiological signal.

2.3 Video Compression and rPPG Algorithm Color Spaces

Algorithms based on first principles such as CHROM and POS have identified the need to focus on the changing color and limit the influence of changes in luminance. To this point in the literature, the connection between this observation for estimating HR and the color spaces used in video compression has not been identified.

2.3.1 POS Projection and YCbCr Color Space

While deriving the POS algorithm, Wang et. al [2] included first-principle motivation for the algorithm steps. This provides an interpretable approach providing a reference for the following comparison. The algorithm input is a time series of the RGB signal extracted from a video. The first step in the algorithm is a projection from RGB to the following basis:

$$B_{pos}^T = \begin{bmatrix} \mathbf{b}_{pos,1}^T \\ \mathbf{b}_{pos,2}^T \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix}, \tag{2.2}$$

where B_{pos} is the basis as written in [2]. While some information from the rPPG signal may be lost in this projection, it has been motivated by a skin reflectance model to retain the cardiovascular pulsatile component while mitigating specular reflection [2]. In the following, we will focus on this projection plane since it is the relevant plane of interest for considering heartbeat signal preservation that can be extracted by POS.

In regards to the impact of compression on the rPPG signal, we will compare the POS projection plane to the Cb-Cr plane from the YCbCr color space defined in Equation 2.1. Consider the

vectors orthogonal to the planes spanned by $[\mathbf{b}_{pos,1}, \mathbf{b}_{pos,2}]$ and the plane spanned by $[\mathbf{b}_{Cb}, \mathbf{b}_{Cr}]$.

$$\mathbf{v}_{pos}^T = (\mathbf{b}_{pos,1} \times \mathbf{b}_{pos,2})^T = \begin{bmatrix} 2 & 2 & 2 \end{bmatrix} \tag{2.3}$$

$$\hat{\mathbf{v}}_{pos}^T = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \tag{2.4}$$

where $\hat{\mathbf{v}}$ indicates a vector normalized to have a magnitude of 1 and $\mathbf{a} \times \mathbf{b}$ is the cross product.

Similarly, we compute and simplify the following using the fact that $1 = K_R + K_G + K_B$.

$$\mathbf{v}_{CbCr} = \mathbf{b}_{Cb} \times \mathbf{b}_{Cr} \tag{2.5}$$

$$= \begin{bmatrix} \frac{1}{4} \frac{K_G K_B}{(1-K_B)(1-K_R)} + \frac{1}{4} \frac{K_G}{1-K_R} \\[2mm] \frac{1}{4} - \frac{1}{4} \frac{K_R K_B}{(1-K_B)(1-K_R)} \\[2mm] \frac{1}{4} \frac{K_R K_G}{(1-K_B)(1-K_R)} + \frac{1}{4} \frac{K_G}{1-K_B} \end{bmatrix} \tag{2.6}$$

$$= \frac{1}{4(1-K_B)(1-K_R)} \begin{bmatrix} K_G K_B + K_G(1-K_B) \\[2mm] (1-K_B)(1-K_R) - K_R K_B \\[2mm] K_R K_G + K_G(1-K_R) \end{bmatrix} \tag{2.7}$$

$$= \frac{1}{4(1-K_B)(1-K_R)} \begin{bmatrix} K_G \\[2mm] 1 - K_B - K_R \\[2mm] K_G \end{bmatrix} \tag{2.8}$$

$$\hat{\mathbf{v}}_{CbCr} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{2.9}$$

Therefore, the projection plane used by the POS algorithm is equal to the Cb-Cr plane in a YCbCr

13

color space. The vector orthogonal to the plane spanned by $[\mathbf{b}_{Cb}, \mathbf{b}_{Cr}]$ is independent of K_R, K_G, and K_B .

2.3.2   Importance of the POS and CbCr Plane Equivalence

The equivalence of the POS projection plane and the CbCr plane that we have discovered above allows for insights related to the preservation of the signal extracted via the POS algorithm when video compression uses the YCbCr color space. The two primary insights are related to spatial down sampling and quantization.

4:2:0 chroma sub-sampling is used for most digital video recordings on consumer electronic devices. In this sub-sampling format, the Cb and Cr color channels are sub-sampled by a factor of two in both the horizontal and vertical directions. For many rPPG applications, a temporal RGB signal is constructed by spatially averaging over a region of interest that contains the subject's skin. Spatial averaging helps overcome quantization noise and average over other uncontrollable factors. In the presence of chroma sub-sampling, this spatial averaging is not averaging over the originally detected values. Instead, the spatial averaging in the Cb-Cr plane is over values that were down-sampled by a total factor of four and then interpolated during video decompression.

The second important factor is related to quantization. It is well known that the human visual system is more sensitive to changes in luminance (Y color channel) and therefore, many video compression algorithms that use a YCbCr color space will limit quantization error in the Y channel at the cost of increased quantization error in the Cb and Cr channels [19]. As a result, a general-purpose video compression program using a YCbCr color space is likely to allocate most bits preserving luminance information that is then discarded by the POS algorithm for rPPG
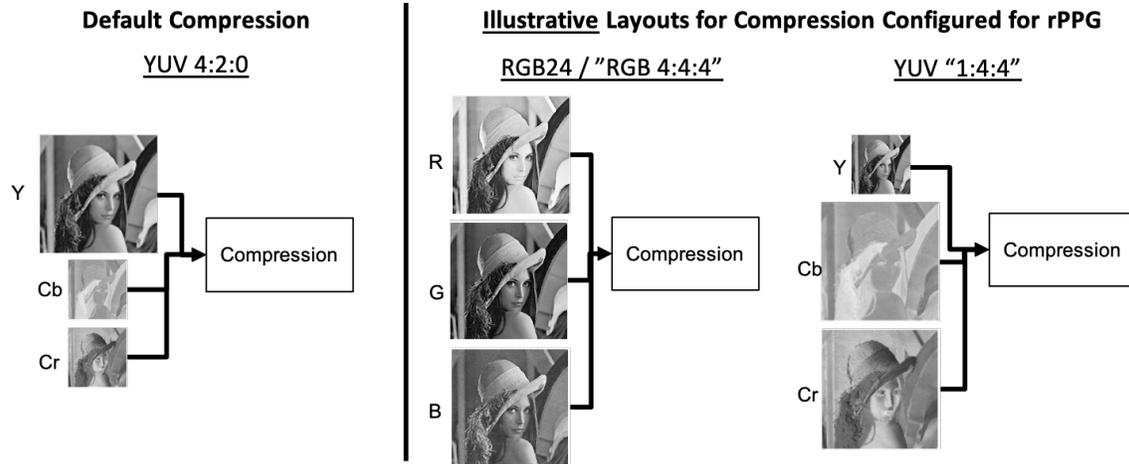
Figure 2.1: Illustrate layouts of video compression configurations that better preserve information leveraged by rPPG algorithms such as POS compared to common default configurations.

analysis. Closely related to video compression standards is JPEG image compression, and it allows for explicit quantization tables that differ for luminance and chrominance [23].

2.4   Application-Specific Video Compression Configuration

Video compression standards such as Advanced Video Coding (AVC) and High Efficiency Video Coding (HEVC) do not impose the usage of a particular color space used during compression. The primary discussion in this chapter is around the YCbCr color space because it is commonly used. However, the coding standards and tools used to compress videos such as FFmpeg, have existing options to change the color space used during compression [24].

Figure 2.1 visualizes illustrative color space layouts that could be used during compression when the application is known to be rPPG. Among these layouts, the RGB24 option currently exists as an option in FFMPEG and is used in the following experiments. The "YUV 1:4:4" layout is not an existing option in FFMPEG because it would include downsampling of the luminance channel which would result in a significant degradation of the visual quality of the video. However, as the

equivalence of the Cb-Cr plane and the POS projection plane showed, this wouldn't impact the POS algorithm when extracting an rPPG signal.

With the increase of remote meetings and the usage of Zoom since 2020, general consumers have become more familiar with the concept of optimizing video compression for a particular task. In the case of Zoom, there is an option to optimize screen sharing for videos. Similarly, there could be a future where there's an option to optimize for physiological monitoring when doing an at-home recording, or during a remote doctor appointment.

2.5   Experiments

After identifying the above relationship between video compression and rPPG algorithms and the opportunity to tune video compression to better preserve the rPPG signal, the following experiments were executed on real and synthetic data to demonstrate the concept.

2.5.1   Simulated Video Signal Preservation

2.5.1.1   Synthetic Data Generation

Synthetic uncompressed videos were produced for experimentation to allow full knowledge of the ground truth time-varying signal in the video. The videos were small, 256x256 videos at 30Hz that each lasted 30 seconds resulting in a total of 900 frames. The function describing the

spatial average of the frame is

$$\begin{bmatrix} r(t) \\ b(t) \\ g(t) \end{bmatrix} = \mathbf{s}(t) = \mathbf{c} + \mathbf{d}A\sin(2\pi ft) \tag{2.10}$$

where $\mathbf{c}$ is the constant color of the frame, $\mathbf{d}$ is a unit vector in the direction of the signal inserted within the RGB color space, A is the amplitude of the inserted sinusoid and f is the frequency of the inserted sinusoid signal.

In the experiments to follow, the signal direction, $\mathbf{d}$, is set in one of two ways. First, $\mathbf{d}_Y^T = 1/\sqrt{3}[1, 1, 1]$ is configured to be orthogonal to the Cb-Cr plane so the signal is only held within the Y channel. Second, $\mathbf{d}_{Cb} \approx [0.00, -0.19, 0.98]$ is configured to be orthogonal to the Y-Cr plane so the signal is only within the Cb channel. In the second case, the ITU-R BT.601 standard is used for setting $K_R = 0.299$, $K_G = 0.587$, and $K_B = 0.114$ [25].

$\mathbf{s}(t)$ represents the spatial average desired in a given frame that will have a non-integer component. The value of a single pixel is set based on a random variable so the spatial average is preserved such that

$$\mathbf{p}_{xy}(t) = \lfloor \mathbf{s}(t) \rfloor + V \tag{2.11}$$

$$V = \begin{cases} 0 & U(0,1) > \mathbf{s}(t) - \lfloor \mathbf{s}(t) \rfloor \\ 1 & U(0,1) \le \mathbf{s}(t) - \lfloor \mathbf{s}(t) \rfloor \end{cases} \tag{2.12}$$

where $\mathbf{p}_{xy}(t)$ is the pixel value at a given position and time, $\lfloor \mathbf{s}(t) \rfloor$ applies the floor operation to each value in the vector, and $U(0,1)$ is a sample from a uniform random variable on the interval

$[0, 1]$. This method results in the expected value of the spatial average of the uncompressed video to be $\mathbf{s}(t)$.

2.5.1.2 Video Compression

The following command was used to compress the video stored in "uncompressed.avi" with the specified parameters.

```
ffmpeg -i uncompressed.avi -c:a copy \
    -c:v libx265 -crf <CRF> -pix_fmt <COLOR_FORMAT> \
    <OUTPUT_FILENAME>
```

In all experiments, the High Efficiency Video Coding (HEVC) standard is used which is also known as H.265. The constant rate factor (CRF) can be set as an integer on the interval $[0, 51]$, where 0 results in lossless compression. The color format was set to either "yuv420p" or "RGB24". The purpose of using yuv420p was to emulate video compression on a consumer electronic device that often uses 4:2:0 chroma sub-sampling by default. The purpose of using the RGB24 setting was to have a comparison of compressed video that does not attempt to leverage attributes of the human visual system.

When changing the color space used for compression, the entropy encoding was not changed which is optimized for YUV420. This means that the resulting bit-rates for the alternative color spaces studied in this analysis may be larger than necessary after re-optimization of entropy encoding for the different color space. Since this optimization would further reduce the bitrate of alternative color spaces, the qualitative results from the analysis are not impacted.

18

2.5.1.3   Feature Estimation

The focus of this work is to illustrate the change of signal preservation with different video compression options and settings. For the first set of experimental results, a small set of results are evaluated and qualitatively compared in the temporal and frequency domain. In the case of the synthetic videos, the full frame was spatially averaged to construct the RGB time series signal.

For the quantitative comparison of signal preservation, the metric for evaluation is the power of the Fourier coefficients at the known sinusoid frequency of 1.217 Hz (73 beats per minute). The purpose of this feature is to consider the preservation of the signal amplitude in different color channels since amplitude preservation has importance for extending rPPG to remote pulse oximetry.

2.5.1.4   Results

Experimental results from synthetic data with known ground truth are presented. The results demonstrate the different behavior in signal preservation of the YUV420 and RGB24 color spaces when the sinusoidal signal is inserted in different directions within the color space. The behavior is inspected in the time and frequency domain for a single CRF and then the behavior is shown for varying CRF. All results align with expectations based on prior knowledge of compression and the color spaces used in the experiment.

Signal Inserted into Y Channel

Synthetic data was generated with the signal inserted orthogonal to the Cb-Cr plane. Recall the YCbCr color space is not orthogonal, so the resulting signal direction is not parallel to the Y
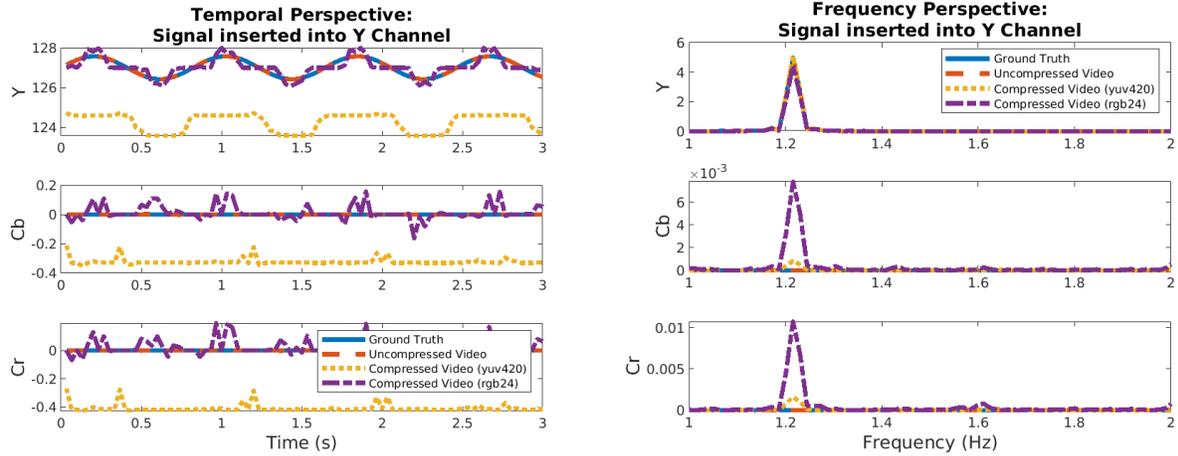
Figure 2.2: Y, Cb, and Cr components in temporal and frequency domain from video constructed with a temporal signal orthogonal to the Cb/Cr plane with a magnitude of 1.

basis vector, however, we will refer to this setup as inserting the signal into the Y Channel. The purpose of this demonstration is to illustrate an example where standard video compression that prioritizes luminance information performs better.

Figure 2.2 shows the results in the temporal and frequency domain when the signal was inserted in the Y channel. The time domain figure visualizes the behavior on a small time window of 3 seconds and the frequency domain figure is computed over the full 30 second window. In this case, the shape of the temporal signal was distorted by video both forms of video compression. However, both video compression methods preserved the power of the signal in the Y channel. The sinusoidal signal is also introduced in the Cb and Cr channels at low amplitudes when the RGB color space is used for compression. This occurs due to the correlation between the RGB color channels introduced by the back-and-forth conversion to and from the YCbCr color space. The exact power of the Fourier coefficient nearest to the true frequency is in Table 2.1.

In the case of compression with YUV420, it is also notable in Figure 2.2 that a bias was introduced to the signal in all three color channels. This work focuses on the effect of compression

Table 2.1: Power in Fourier transform coefficient nearest the inserted signal frequency of 1.217 Hz (73 beats per minute).

Color Channel	Source	Signal In Y Channel		Signal in Cb Channel	
		Magnitude	Error	Magnitude	Error
Y	Ground Truth	4.991	n/a	0.000	n/a
	Uncompressed	4.991	0.000	0.000	0.000
	Compressed (YUV420)	5.122	0.131	4.105	4.105
	Compressed (RGB24)	4.444	-0.547	0.215	0.215
Cb	Ground Truth	0.000	n/a	4.596	n/a
	Uncompressed	0.000	0.000	4.596	0.000
	Compressed (YUV420)	0.001	0.001	1.537	-3.059
	Compressed (RGB24)	0.002	0.002	4.433	-0.163
Cr	Ground Truth	0.000	n/a	0.000	n/a
	Uncompressed	0.000	0.000	0.000	0.000
	Compressed (YUV420)	0.002	0.002	0.072	0.072
	Compressed (RGB24)	0.011	0.011	0.110	0.110

on small time-varying signals so this effect is not investigated further. However, this bias may have effects due to error being introduced into the denominator of the RoR method if a remote pulse oximetry methodology uses the same approach.

A similar set of experiments was evaluated with the same uncompressed videos but now varying the CRF for both compression schemes. Both YUV420 and RGB24 preserved the signal relatively well for low CRF below 20. Beyond a CRF of 20, the RGB24 videos have significantly more loss of signal preservation relative to YUV420 which maintains good performance until the CRF exceeds 37 (Figure 2.3). This aligns with the expected behavior based on knowledge of compression schemes using the YUV color space.

Signal Inserted into Cb Channel

The same experiment was evaluated with the signal inserted in the direction orthogonal to the Y-Cr plane which will be referred to signal in the Cb channel. Figure 2.4 demonstrates the
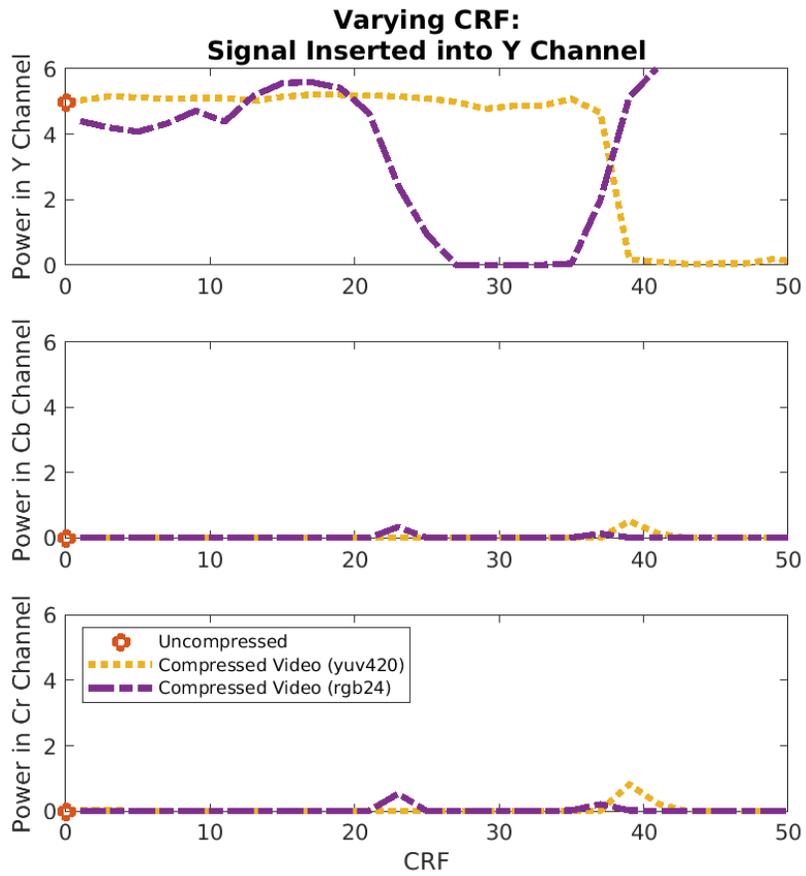
Figure 2.3: Behavior of varying CRF for video constructed with a temporal signal orthogonal to the CbCr plane with a magnitude of 1.

behavior at a single CRF in the temporal and frequency domain.

This demonstration illustrates behavior when the signal is inserted into a color channel which is important for rPPG signal extraction algorithms such as POS. As before, the shape of the temporal signal changes for both video compression methods. However, the signal magnitude preservation between the YUV420 and RGB24 compression results is significantly different. The signal amplitude in the Cb color channel is significantly reduced with YUV420. In the case of RGB24, the amplitude in the Cb color channel is preserved but some signal is similarly introduced into the Y and Cr channels. Again, this correlation occurs due to compression occurring in a different color space than what is being plotted. As before, the exact powers at the true signal frequency are in Table 2.1.

The difference in signal preservation between the two cases highlights the fact that the YUV420 video compression approach was developed with the goal of visually pleasing compression results for humans. Stronger human sensitivity to the Y color and reduced sensitivity to Cb and Cr has resulted in less ability to preserve a temporal sinusoid when the signal was inserted orthogonal to the Y-Cr plane. In the case of heart rate (HR) estimation with rPPG, the amplitude distortion has not yet been a significant problem due to the information being in the existence of the HR signal rather than the magnitude. However, for SpO_2, the information is in the amplitude of the rPPG signal in different color channels making this of higher importance.

As before, the same methods were evaluated for varying CRF (Figure 2.5). The YUV420 videos never preserve the Cb signal well even at low CRF. The RGB24 videos preserve the Cb signal relatively well until the CRF exceeds 20 similar to the behavior when the signal was in the Y channel.
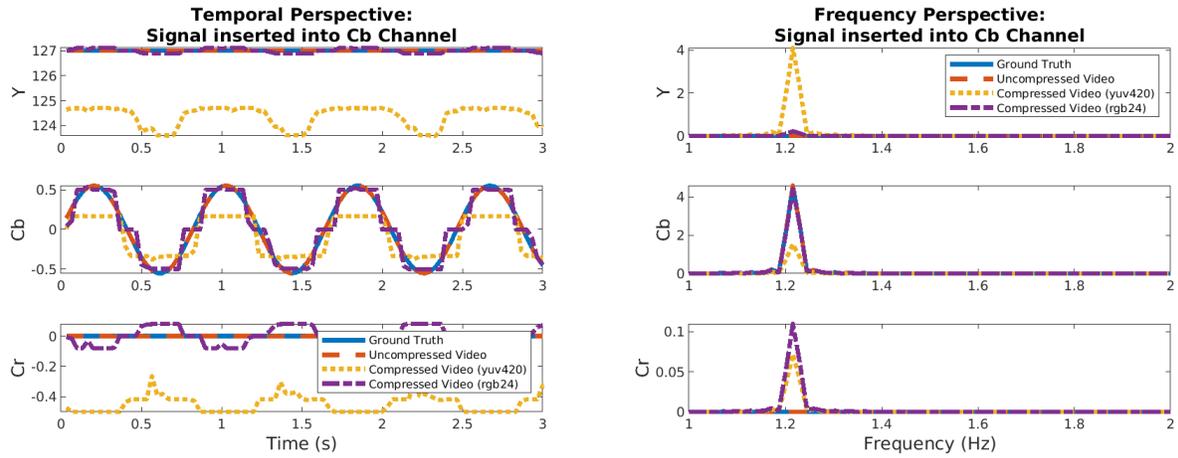
Figure 2.4: Y, Cb, and Cr components in temporal and frequency domain from video constructed with a temporal signal orthogonal to the Y/Cr plane with a magnitude of 1.
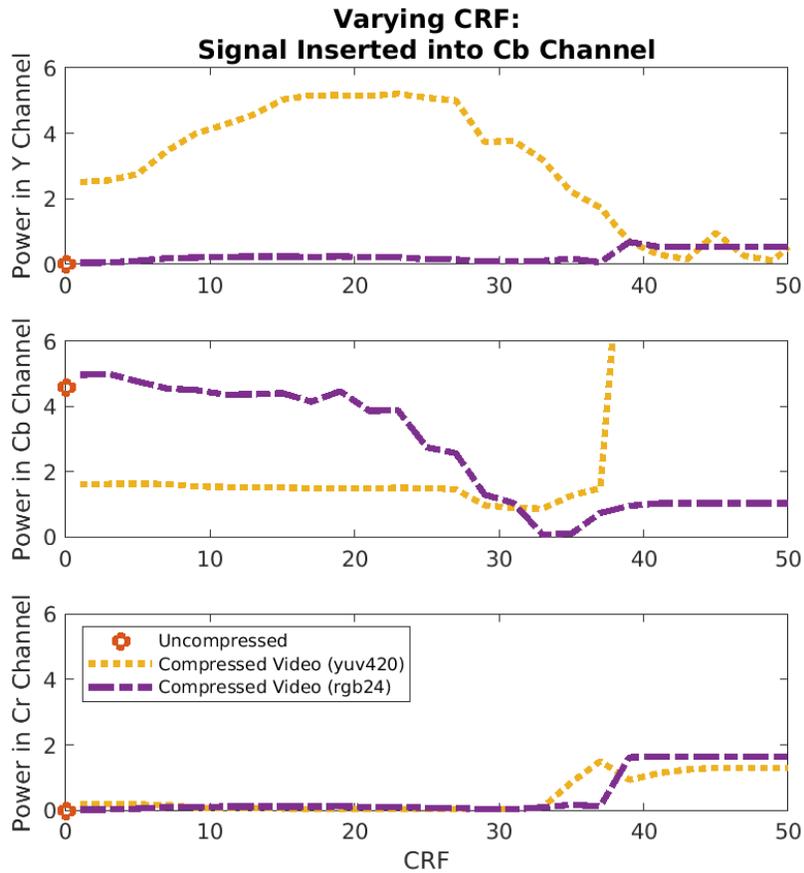

Figure 2.5: Behavior of varying CRF for video constructed with a temporal signal orthogonal to the Y/Cr plane with a magnitude of 1.

24

2.5.2 Heart Rate Signal Preservation

After successfully demonstrating the impact that the video compression color space can have on small signal preservation for synthetic data samples, an experiment was designed to illustrate the behavior on HR data collected by a camera. The data collection methodology was intentionally configured to result in a high amplitude rPPG signal to create clear results. The data analysis focused on the preservation of amplitude at the HR frequency with an emphasis on the red color channel due to the highest amplitude caused by the collection method.

2.5.2.1 Data Collection

For data collection, the study protocol was approved by the University of Maryland Institutional Review Board (protocol #1376735-2). Informed consent was obtained from each participant.

Data was collected with a Sony RX 10 IV camera. This camera model was selected due to its ability to collect high frame rate JPEG images at approximately 20 Hz for about 10 seconds. The sequence of JPEG images with a quality level of 100 was considered to be the "uncompressed" signal for the following experiments. The images did contain spatial compression, but no temporal compression, which is the focus of this work.

High amplitude rPPG signal was enabled by recording a subject's finger backlit by a smartphone flashlight. Ground truth recording of the subject's heart rate was collected with a pulse oximeter on the left forefinger and the rPPG signal was collected from the right forefinger. A picture of the experimental setup is shown in Figure 2.6.

An example frame from a collection with a manually selected region of interest is shown in Figure 2.7. The camera was manually configured to have a constant exposure time and ISO settings

Figure 2.6: Experimental setup with ground truth recording on the left forefinger, a smartphone light behind the right forefinger, and the camera collecting data over a plain background.
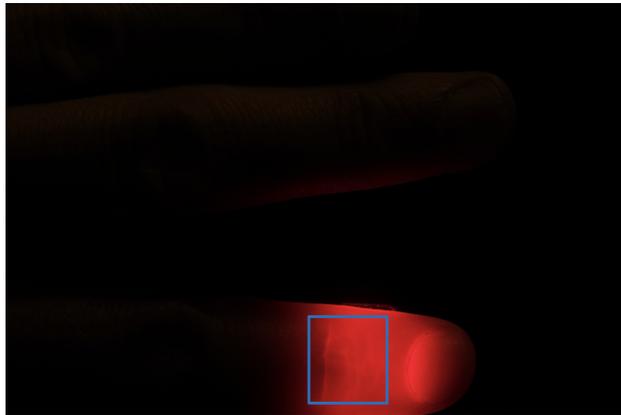
Figure 2.7: An example frame with a manually selected region of interest.

during each data collection for consistency during a frame sequence and to avoid saturation.

The same process was used for applying video compression with FFMPEG. For this test, the color space option YUV444 was added to the previous test that only included RGB24 and YUV420. YUV444 does not include spatial downsampling and allows for a comparison without that additional factor to consider when interpreting results.
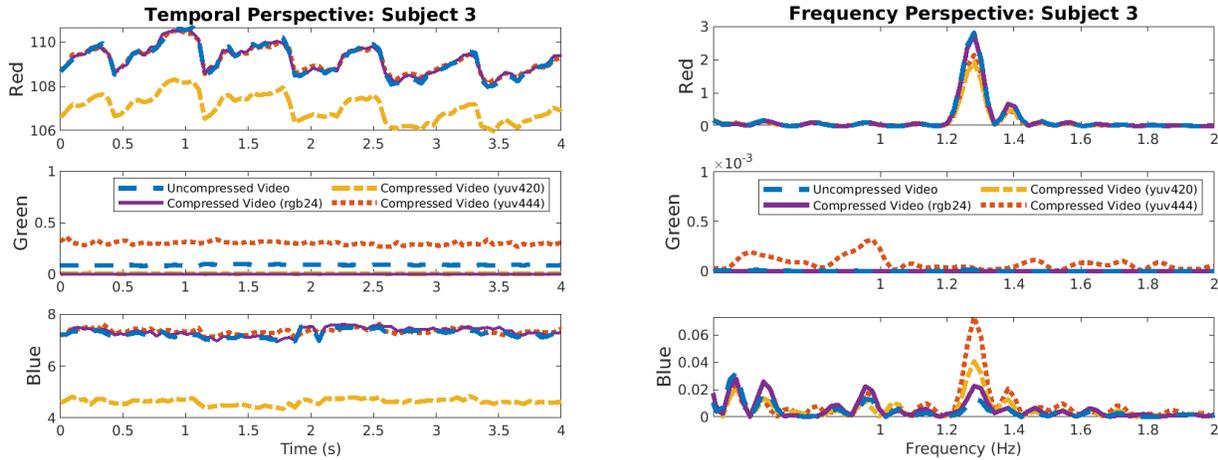
Figure 2.8: Example of RGB signals extracted from a manually selected region of interest in the time and frequency domain.

2.5.2.2 Results

Figure 2.8 shows a representative example with the temporal and frequency perspectives at a single CRF of 23 (bitrate $\approx$ 0.25 Mb/s). This particular CRF is shown to illustrate the behavior in the regime of bitrates where there is different behavior among the compression configurations. For this analysis, the red rPPG signal is the primary focus due to the strong rPPG signal. The green and blue channels have minimal HR signal due to low DC components at the specific exposure setting. This is because the Bayer color filter array applies equal exposure to all color channels, and increasing the exposure to have more signal in the green and blue channels would have caused over-saturation for skin areas in the red channel. A qualitative observation directly related to the usage of the YCbCr color space during compression is the increase in power at the HR frequency in the blue channel for compressed video. This occurs due to the correlation between the RGB color channels introduced by the back-and-forth conversion to and from the YCbCr color space during video compression.

27

Figure 2.9: Estimated power of the HR signal amplitude in the red color channel as a function of bitrates when different color space options were used during compression from a single subject.
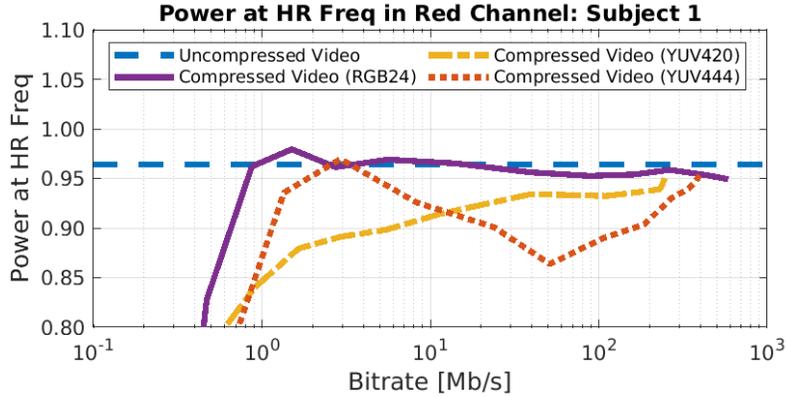
A parameter sweep over the CRF and the investigated color space options (RGB24, YUV420, and YUV444) produced the videos used to extract power at the known HR frequency for each configuration. The HR recorded by the pulse oximeter was used to identify the relevant frequency for each subject in the 10 second window. The estimated power of the HR frequency as a function of bitrate is shown for subject 1 in Figure 2.9 with a comparison to the output in the uncompressed video (5 Gb/s). The results analogous to Figure 2.9 were aggregated across the four subjects and the minimum, mean, and maximum are plotted in Figure 2.10. The minimum and maximum bounds are shown in favor of the standard deviation due to the small sample size used for the concept demonstration in this thesis.

Reading Figure 2.10 from right to left, the compressed videos at high bit rates have small relative errors. As the curve moves to the left, indicating lower bitrates, the estimated power diverges from the reference. The videos that used the RGB24 color space maintained lower relative error of HR amplitude estimation in the bitrate range from 1-20 Mb/s for most cases. Similar qualitative behavior of the relative performance of different color spaces was also observed when the ground truth power (blue reference line in Figure 2.9) varied from 0.2 to 2.8. These results align with the previously described expectations based on knowledge of the YCbCr/YUV color

28
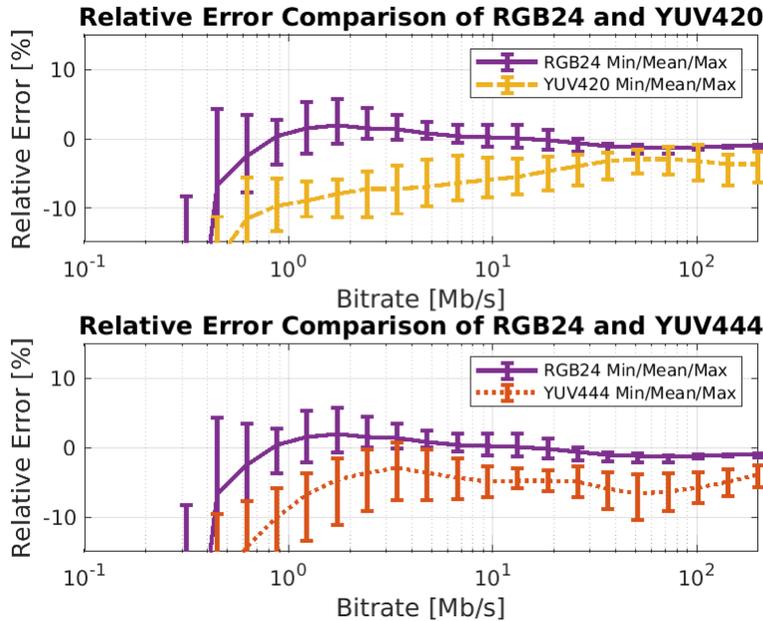
Figure 2.10: Minimum, mean, and maximum relative error as a function of bitrate aggregated over the four subjects plotted for RGB24/YUV420 and RGB24/YUV444 comparisons.

space and the CbCr spatial downsampling.

The YUV444 configuration uses YCbCr during compression, but does not spatially downsample the chrominance channels. This configuration is not commonly used for compression, but the result that RGB24 had better performance than YUV444 at lower bitrates demonstrated the impact the color space has on the rPPG signal preservation in the absence of spatial downsampling. The relative performance of YUV444 and YUV420 was not always consistent across all of the subjects. This is because at a specific bitrate, attributes of a video may impact the dynamic behavior of the HEVC coding, resulting in better performance with or without subsampling of chrominance components. Due to the subtle difference between YUV420 and YUV444, a larger sample size would be required to draw strong conclusions, but that direct comparison is not relevant to the conclusion of this work. Rather, the result that RGB24 had better performance than YUV both with and without spatial downsampling is the predicted result these data confirmed.

Preserving the amplitude of HR signals is particularly important for SpO_2 estimation applications because the information related to blood oxygen saturation is in the amplitude of the signal [26]. When building a robust system for SpO_2 estimation, it is important to consider amplitude error introduced by video compression to be an additional error source that is added to other modeling and approximation errors for SpO_2 estimation with consumer electronic devices. Improved preservation of the HR signal can also benefit HR detection in other physiological applications.

2.6   Compression and Remote Pulse Oximetry

To relate the relative error results back to SpO_2 estimation, consider the case of using the RoR method with red and green similar to [8]. The ratio of ratios in this case is

$$RoR = \frac{AC_r/DC_r}{AC_g/DC_g} = \frac{AC_r}{AC_g}\frac{DC_g}{DC_r}. \tag{2.13}$$

If we assume the DC components do not have error, then the relative error of the RoR is fully determined by the errors in the AC components. For illustrative purposes, Figure 2.11 shows the relative error of the RoR as a function of the relative error of AC_r and AC_g. When the relative error is similar in both color channels, the error cancels out due to the ratio, but different errors in the two color channels results in an inaccurate estimate of the RoR and consequently the SpO_2 estimate.

For additional context, the relative error in RoR to the error in estimated SpO_2 is mapped using the following fit from Moco et al. [8]

$$SpO_2 = -93.1RoR + 104.6 \tag{2.14}$$

30
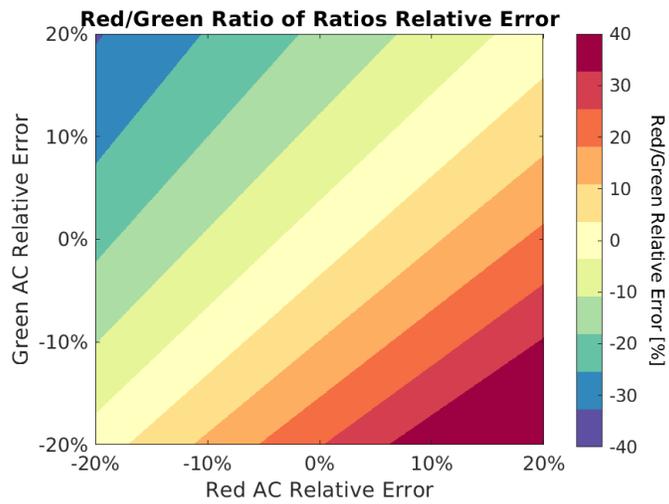
Figure 2.11: Red/green ratio of ratios relative error as a function of relative error in the AC component in the red and green channels individually.
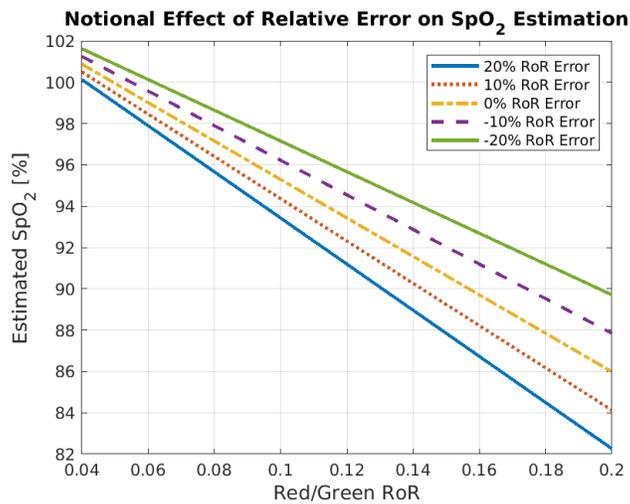
Figure 2.12: Notional effect of relative error of RoR on estimated SpO_2 using linear fit from Moco et al. [8]

Figure 2.12 plots the estimated SpO_2 using this fit at different levels of RoR relative error. The maximum absolute error for SpO_2 estimation in pulse oximeter ISO requirements is 4% [27]. As previously mentioned, there are approximations that result in model error when extending traditional methods for SpO_2 estimation to use consumer cameras. This example is to illustrate the additional estimation error that may be introduced from compression.

Chapter 3

Multi-Spectral Sensor for Physiological Monitoring

3.1 Motivation & Purpose

The prior chapter focused on remote physiological monitoring with common consumer RGB cameras. This line of work has the potential to benefit a large population due to the widespread availability of these devices, however, there are a variety of previously mentioned limitations such as broad spectral responsivity and identical exposure settings in all color channels. In the case of broad spectral responsivity and ambient light, calibration becomes difficult because changes in value can come from the subject or changes in the ambient light spectrum. This makes it difficult to calibrate due to the integral of products visualized in Figure 1.5.

Early work extending remote HR monitoring to demonstrate the feasibility of remote pulse oximetry had experimental configurations that kept information in the recorded videos specific to a relatively narrow spectral band. Verkruysse et al. demonstrated the feasibility of calibration using a custom setup with monochromatic cameras and relatively narrow pass-bands in the red and NIR [7]. Moco et al. extended the work and applied a similar methodology to show feasibility with similar pass-bands in red and green bands [8]. As an alternative to having multiple cameras with different optical filters, Shao et al. configured a customized lighting configuration to enable

recordings with more spectral specificity. In their setup, they had a panel of LED lights at 611 and 880 nm that were alternated in combination with the camera trigger so every other frame had alternate lighting [3]. Further studies have demonstrated remote pulse oximetry with ambient light and RGB sensors, but the physical model indicates that avoiding these approximations should lead to a more accurate system. These studies approach the problem of achieving spectral specificity in different ways by either manipulating $T_O(\lambda)$ with optical filters, or $L_r(\lambda)$ with custom lighting (Figure 1.5).

In these research studies, customized sensor setups or lighting configurations were required to demonstrate feasibility. One direction of research stemming from this body of work is toward standard consumer cameras with a focus on what is available in smartphones due to broad availability and applicability. Another attractive direction of future work is evaluating the capability of commercial off the shelf (COTS) sensors that retain spectral specificity without requiring customized setups. This direction of work is more likely to be applicable in clinical or commercial settings where it is worthwhile to have a custom sensor for remote physiological monitoring if it results in more accurate estimations.

The Sentera 6x sensor is a candidate COTS sensor for this future research. In this chapter, attributes and observations of the Sentera sensor are identified and investigated to understand the impact to the rPPG signal and implications applicable to future work. The primary observations are related to the 5 Hz sampling rate of the camera, the automatic exposure controller built-in to the camera, inconsistent exposure from the RGB camera, and necessary steps towards efficient and effective multi-spectral video processing.

Figure 3.1: Sentera 6x camera and visualization of image slices offered by the sensor. Images from https://sentera.com/data-capture/6x-multispectral/.

3.2 Sentera 6x Sensor Overview

Sentera is a company that develops sensors with the primary targeted application of smart agriculture enabling the evaluation of crops from sensors aboard drones. In the following work, the Sentera 6x Sensor (Figure 3.1) was tested due to the narrow spectral bands and the ability to record at 5 Hz.

The cameras included in the Sentera 6x include 5 monochromatic bands in the blue, green, red, red-edge, and near-infrared regions of the spectrum. Table 3.1 outlines a comparison of the available bandwidths with the optical setup from [7, 8]. In all cases, the Sentera 6x sensor offers a narrower bandwidth compared to these previous studies, so the study characteristics should apply from a spectral responsivity perspective. In addition to the monochromatic cameras, the 6x sensor also has an RGB camera with a Sony IMX147 detector with a rolling shutter.

When operating in continuous capture mode, the sensor saves directories for each monochromatic channel and the RGB camera. For the monochromatic channels, each frame is stored in a

35

Table 3.1: Comparison of optical filters from [7, 8] and Sentera 6x sensor.

Color	From Verkruysse et al. and Moco et al.		Sentera 6x	
	Center Wavelength [nm]	Bandwidth [nm]	Center Wavelength [nm]	Bandwidth [nm]
Blue	N/A	N/A	475	30
Green	579	32	550	20
Red	675	67	670	30
Red Edge	N/A	N/A	715	10
NIR	842	56	840	20
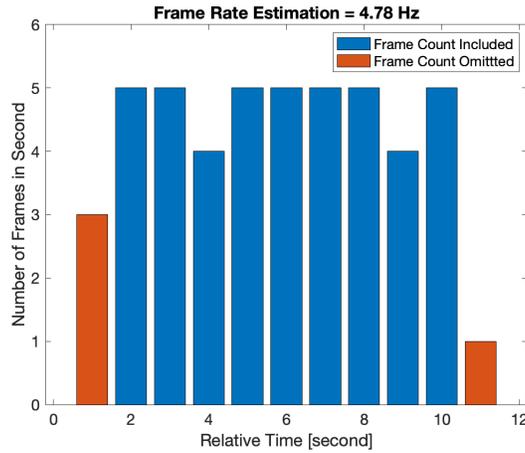
Figure 3.2: Example of fractional frame rate estimation when the timing information is only stored with one second of precision. The frequency estimation is the average of the number of frames in each second after omitting the first and last seconds.

Tagged Image File Format (TIFF) file with 12 bits of precision for each pixel and the images are 2048 x 1536. These files also include metadata associated with the camera, timing, and exposure settings. The RGB camera also saves individual files in a JPEG format with 8-bit precision and a quality setting of 75% to allow for compression of the 5184 x 3888 frame.

The timing information stored in the TIFF and JPEG files is only recorded with 1 second of precision. The frame rate was found to not be precisely 5 Hz. The fractional frame rate was estimated by computing the average number of frames recorded in each second after omitting the first and last second as visualized in Figure 3.2.

36
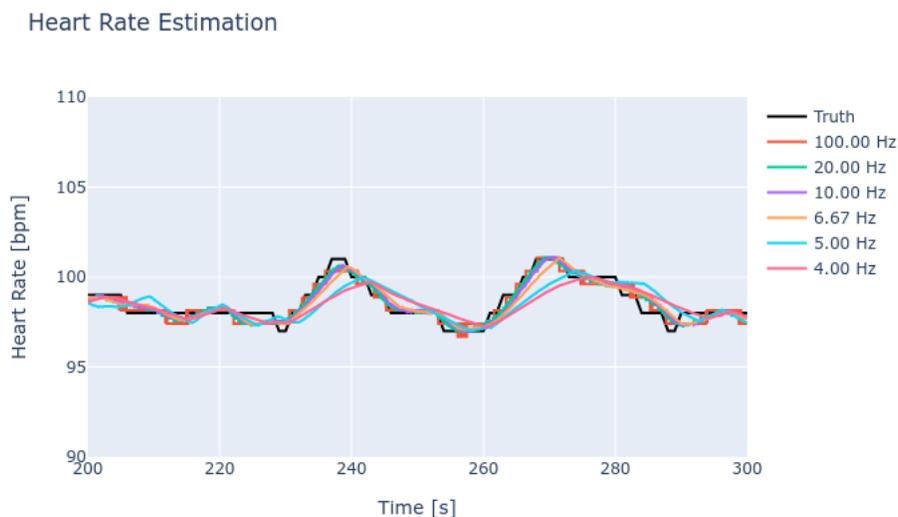
Figure 3.3: Heart rate estimation on downsampled data from the PhysioNet dataset.

3.3 Feasibility of 5 Hz HR Monitoring

The spectral characteristics of the Sentera 6x is the primary feature that makes it a good candidate for rPPG applications. One negative attribute of the camera for this application is it is only capable of recording at 5 Hz when operating in continuous capture mode. This device limitation is a characteristic of the primary application being smart agriculture where the spectral information of interest does not vary quickly with time.

The Nyquist rate of 5 Hz indicates the maximum detectable heart rate without aliasing is 150 beats per minute (BPM). In practice, the upper limit is likely lower due to the PPG signal not being a pure sinusoid and containing some frequencies above the HR frequency. To demonstrate HR estimation at low sample rates, a test PPG signal from the PhysioNet dataset [28] was downsampled without an anti-aliasing filter to model an under sampled sensor. After downsampling, Adaptive Multi-Trace Carving (AMTC) evaluated and the results are shown in Figure 3.3. AMTC is a unified

**SNR of Sinusoid with Amplitude 1
in WGN with σ^2=1 in 10 s window**

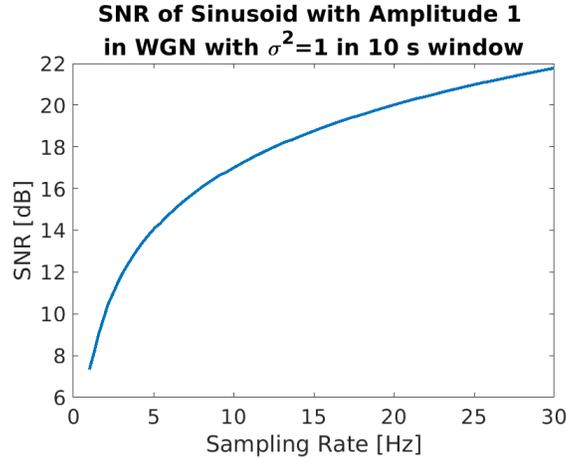Figure 3.4: Demonstration of the effect of the sampling rate on the SNR of a sinusoid with an amplitude of 1 in additive white Gaussian noise (WGN) with variance 1 for a 10 second window.

approach to the detection and estimation of low SNR signals developed by Zhu et al. [29]. The results align with the expectations based on the Nyquist rate.

One notable attribute of the 150 BPM upper limit when using the Sentera 6x is experiments and applications should be limited to non-exercise scenarios. Remote physiological monitoring during physical activity is an area of previous work [30], but would not be applicable with the Sentera 6x.

Evaluation of the Nyquist rate considers the potential impact of aliasing, but another consideration is the impact of noise. When computing the Discrete Fourier Transform (DFT), the expected level of the white Gaussian noise will not depend on the number of samples. However, the power in the DFT at the frequency of a sinusoid will depend on the total power in the sampled signal, which varies with the sampling rate. If you consider an example with a fixed 10 second observation window of a sinusoid with an amplitude of 1 in additive white Gaussian noise with a variance of 1, you can evaluate the SNR as a function of variance as shown in Figure 3.4. In this example, there is approximately an 8 dB difference in the SNR when you compare 5 Hz to

30 Hz. This is an example based on the same fundamental sigal processing principles from the well known fact that increasing sampling rate will increase SNR [31]. This behavior can have a significant impact and indicates collections with the Sentera 6x will be much more sensitive to noise due to the lower collection frame rate.

3.4 Automatic Exposure Controller Observations

In many cases, consumer cameras record with constant exposure settings, but this was found not to be the case with the Sentera camera. Additionally, the exposure settings of each color channel are individually controlled. This has important implications for the usage of data coming out of the different color channels because

$$L \propto \frac{DN}{t_{exposure} * ISO} \, ,$$

(3.1)

where L is the incident radiance at the aperture, DN is the digital counts recorded by the sensor, $t_{exposure}$ is the exposure time, and ISO is a gain factor. An example of the varying exposure time and gain during a single collection is shown in Figure 3.5.

When evaluating a time series that was collected with dynamic exposure, it is important to properly normalize for exposure time and ISO so the time series is proportional to the radiance at the aperture and not dependent on sensor configuration. This is similar to operating on calibrated radiometric imagery from electro-optical and infrared scientific equipment. Additionally, if information from multiple color channels is being used for a future study and the relative values are of interest, then proper normalization for exposure time and ISO gain will be important.

Experimentation also suggested the internal exposure controller targets a relatively naive
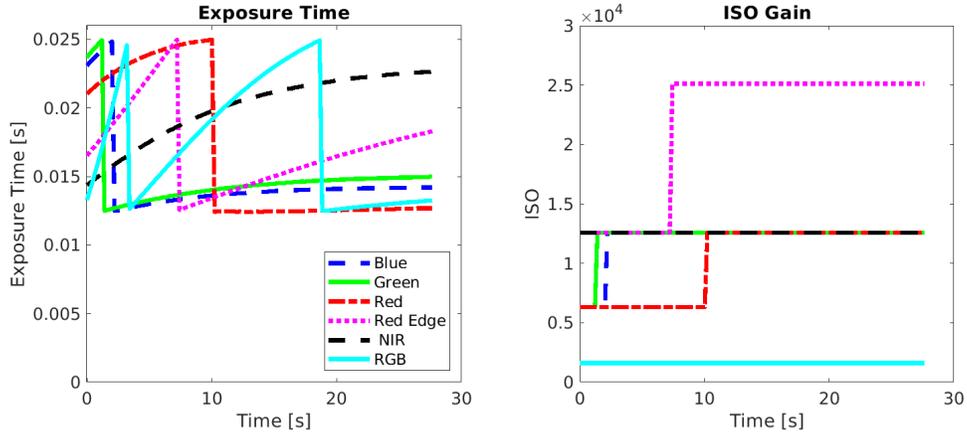
Figure 3.5: Example of time-varying exposure time and ISO gain at beginning of collection with Sentera 6x Sensor.

Table 3.2: Exposure settings and mean counts in direct and indirect sunlight demonstrating nearly identical average counts in all color channels.

Color	Indirect Sunlight			Direct Sunlight		
	Exposure Time [ms]	ISO	Mean Counts	Exposure Time [ms]	ISO	Mean Counts
RGB	2.63	1600	92	1.00	200	93
Blue	2.40	1585	1542	1.18	197	1576
Green	2.81	1525	1525	1.05	197	1540
Red	1.71	1585	1518	1.00	100	1547
Red Edge	3.57	1585	1534	1.36	197	1571
Near Infrared	2.65	1585	1523	1.38	197	1547

state of a constant average value in the frame. Example exposure settings and resulting mean counts for two collections from indirect and direct sunlight are shown in Table 3.2. This naive exposure controller resulted in overexposure and saturation of light-skin pixels when the background was dark. When designing experiments with the Sentera 6X in the future, the similarity of background and foreground will be important to avoid this over-saturation.

3.5   Observations of Inconsistent RGB Exposure

The automatic exposure controller for the cameras was a primary observation during initial experimentation. An additional step was taken to evaluate the consistency of the recorded exposure settings during a collection and the correlation between the reported exposure time and the average value of the collected frame. Collections were executed with the cameras looking at a flat background of cardboard. There are some low frequency time variations of the signals, but the focus of the following is the high frequency behavior. In the following results, the ISO gain was constant.

First, Figure 3.6 shows the spatial average of the monochromatic cameras and the corresponding exposure times. The exposure time controller for the Sentera 6x appears to be very sensitive and commonly switches between exposure times. The device specifications do not report the precision with which exposure time can be controlled for the monochromatic cameras, but the results appear to indicate the controller is often flipping between adjacent exposure times within the discrete set of options. This result emphasizes the prior point that the time series constructed from the Sentera 6x sensor needs to be appropriately normalized by exposure time and ISO gain.

Next, Figure 3.7 shows the same spatial averages from the RGB camera when observing a piece of cardboard. For the RGB camera, the exposure time is much more consistent compared to the monochromatic sensors. However, there is an additional observation that there is step-like behavior in the resulting spatial averages indicating variation between two exposure settings. This behavior was investigated, but no solution was found. Extracting low SNR rPPG data in a time series with the additional non-Gaussian noise caused by this behavior will be difficult.
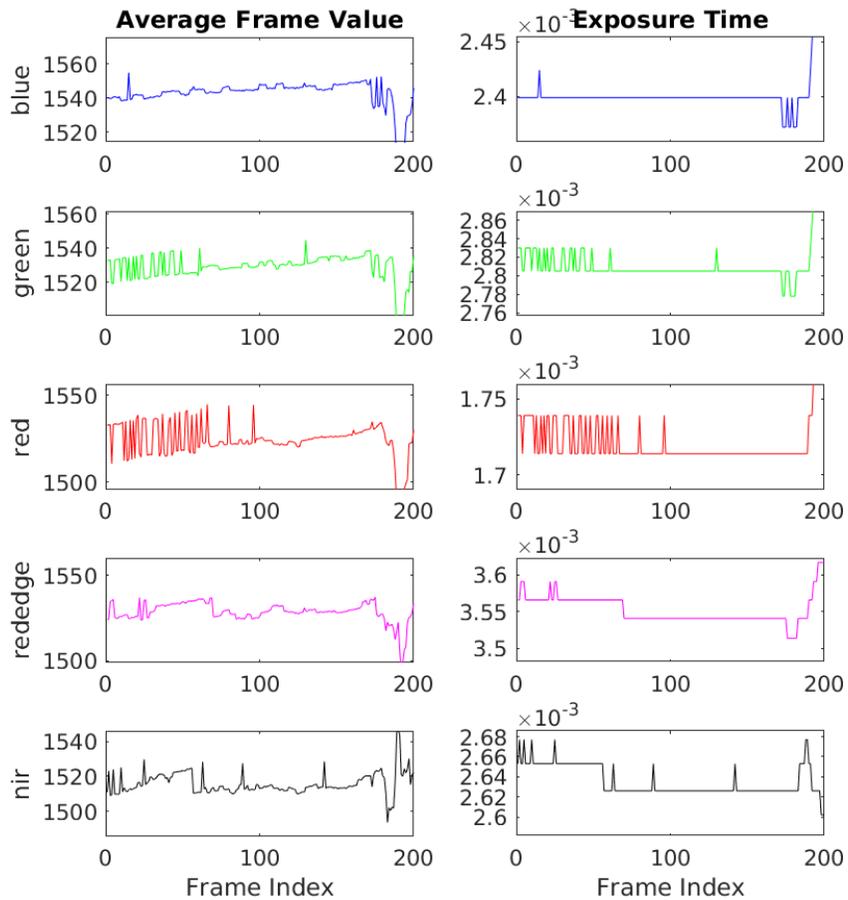
41

Figure 3.6: Spatial average of monochromatic cameras and corresponding exposure times for collection with a near constant background of cardboard.
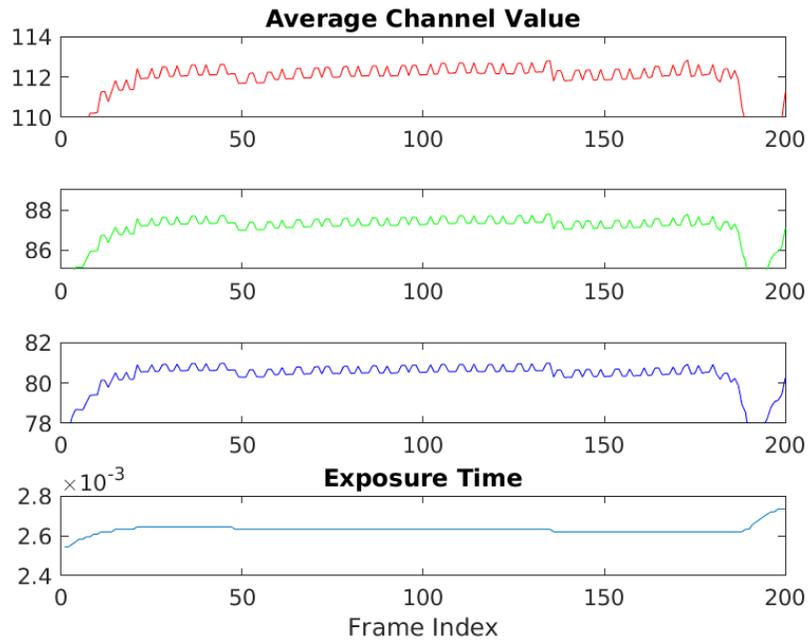
Figure 3.7: Spatial average of RGB camera and corresponding exposure time for collection with a near constant background of cardboard.

3.6 Frame Registration Demonstration

The primary application under consideration during initial testing was remote physiological monitoring extracted from videos of the face. To execute this task, a step in the process for the algorithm pipeline would need to be face recognition and skin segmentation. In the case of the Sentera sensor, there are six video streams and it would be undesirable, inefficient, and error-prone to individually evaluate face recognition and skin segmentation algorithms in all six video streams independently.

Ideally, facial recognition and skin segmentation would be executed from the RGB camera. This preference is not due to any specific physical information available in the RGB video stream, but rather the portability of existing packages and tools such as DLIB [32] that leverage existing

43

Figure 3.8: Example of frame registration across all color channels enabling repeated use of face detection and skin masking.

algorithms and research to avoid replication of face detection.

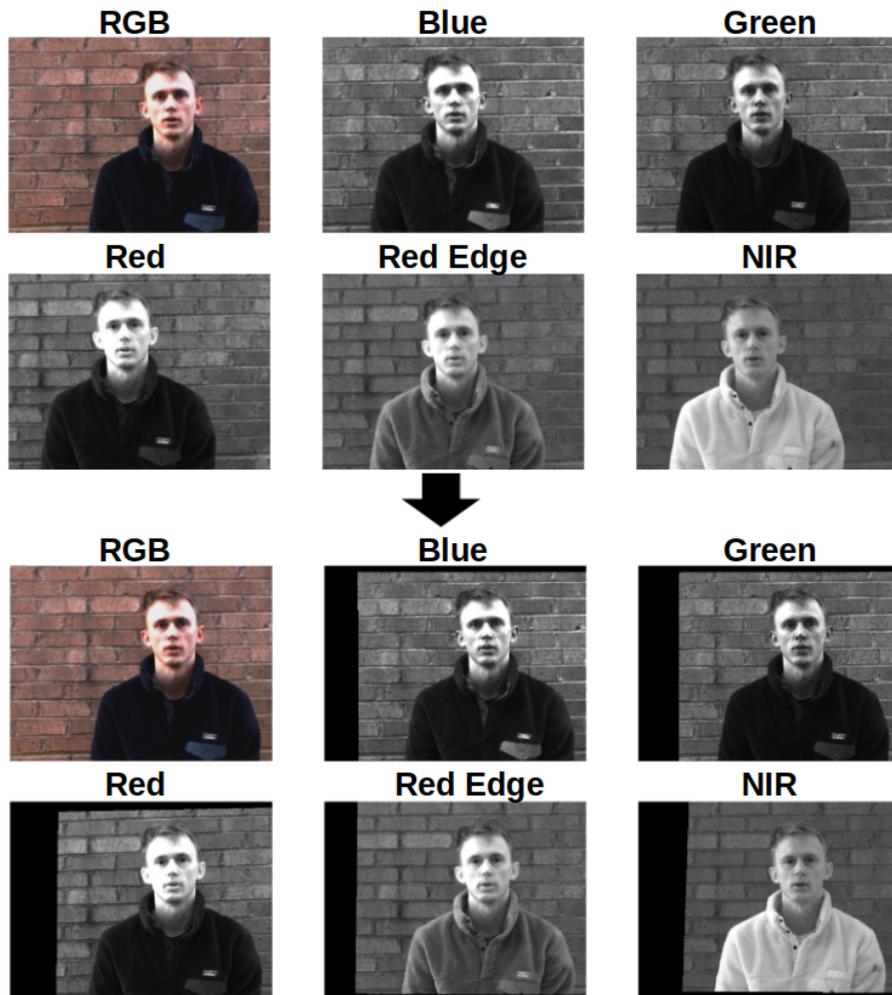Under a controlled setting with a known distance between the camera and the subject, simple calculations could be used to compute the displacement between the subject in the RGB frame and each of the monochromatic cameras after appropriate calibration. Due to the static relative position and pointing of the cameras in the Sentera 6x, the computation would be based on standard stereo camera geometry. However, this method would not be sufficient in a more likely scenario of unknown distance to the subject.

A more adaptive approach would actively estimate frame registration parameters between the different color channels. In practice, this computation would likely not happen on every frame, but rather periodically to re-calibrate the alignment. The following is a concept demonstration of frame registration across the various color channels. This demonstrates the applicability of the standard frame registration procedure of key point detection, key point matching, and homography estimation.

To produce the results in Figure 3.8, open-source tools available in OpenCV were used and the specific algorithms for each step are as follows.

- Key Point Detection: ORB Feature Extraction [33]

- Key-Point Matching: Brute Force

- Affine Estimation: RANSAC [34]

The results successfully demonstrated the feasibility of using the standard approach for the registration of images in different wavebands with visibly different features in terms of absolute intensity. This result aligns with the fact that ORB features (and similarly for SIFT and SURF features) rely most heavily on gradients and edges in the image. For a more specific experiment

45

with Sentera 6x, the algorithm selection for the standard registration pipeline could be refined particularly for the selection of brute force key-point matching.

3.7   Summary

In this chapter, key observations about the characteristics of the Sentera 6x hardware and software behavior were identified with a primary focus on applications for remote physiological monitoring. Some of these characteristics are caused by the fact the sensor has been developed with the target application of smart agriculture. The previously presented observations and implications are summarized in Table 3.3.

Overall, this sensor has the attractive feature of being a COTS multi-spectral sensor that could be suitable for remote physiological monitoring using existing devices. The direct incorporation of the sensor in future research studies and analysis for physiological monitoring will face challenges identified in this chapter's investigation stemming from the sensor properties designed for the different purpose of remote agricultural sensing. However, these issues can be overcome with the appropriate measures from the algorithms and the implementation of signal analysis.

Table 3.3: Summary of contributions to enable future work with Sentera 6x sensor.

Attribution/ Observation	Direct Impact	Implication
5 Hz Sampling Rate	Nyquist rate of 150 bpm	Exercise heart rate causes aliasing
5 Hz Sampling Rate	Decreased SNR in Frequency Domain	Low SNR will increase data processing challenges
Automatic Exposure Controller	Direct time series of counts is not proportional to incident irradiance	Data processing needs to appropriately normalize for exposure time and ISO gain
Automatic Exposure Controller	Monochromatic videos appear to target average counts of $\sim$1550	Experimental setup needs to ensure background is similar enough to skin tone to avoid skin saturation
Inconsistent RGB exposure	Average values from RGB camera vary significantly under constant exposure settings	Direct use of RGB stream from Sentera 6x sensor may not be usable for rPPG
Multi-Video Stream Processing	Some algorithmic steps such as region-of-interest selection need to be applied to multiple video streams	Options such as frame registration can be used to leverage existing processing of RGB videos

Chapter 4

Summary and Future Perspectives

This thesis has contributed to the research area of remote physiological monitoring with a focus on the sensing step in the process (Figure 1.1). This focus has resulted in new identification of the significant relationship between video compression and the rPPG signal and characterization of the Sentera 6x sensor to assist usage in future studies that will leverage the precise multi-spectral precision compared to standard RGB cameras.

4.1 Video Compression and Cardiac Monitoring

Chapter 2 focused on the close relationship between the YCbCr color space commonly used in video compression and the information leveraged by rPPG algorithms, such as POS, that are based on first principles. This identification motivated an extension of existing work in the literature that has evaluated the SNR of the rPPG signal at varying compression levels. This observation motivated the question of if the color space used for video compression could be selected to better preserve the rPPG signal. Fortunately, configuration options already exist in video compression codecs such as HEVC which enabled experimentation on low amplitude signal preservation in synthetic and collected data.

First, synthetic videos were generated with precisely known ground truth of the signal. A

48

sinusoidal signal was inserted into the video sequence such that the time series constructed with spatial averages of the frame would create a low amplitude sinusoid. This experimental setup was intended to emulate low amplitude rPPG signals. The original uncompressed video was compressed using FFMPEG at varying CRF resulting in different bit rates. The experiment was executed with video compression configured to use the YUV420 color space and the RGB24 color space. Additionally, the experiment was executed when the sinusoid was intentionally inserted into the Y channel and the Cb channel. As expected from video compression principles and configurations, the YUV420 configuration was able to preserve the Y channel sinusoid at lower bit rates compared to the videos compressed using the RGB24 color space. And conversely, the RGB24 configuration better preserved the sinusoid when it was inserted into the Cb color channel. Successful correlation between expectation and experimental results in the controlled environment of synthetic data motivated the next step with real sensor data.

The goal of the experiment was to demonstrate varied behavior of preservation of the HR signal using different color spaces for video compression. This goal required the capability to record uncompressed data. The Sony RX 10 IV was able to collect about 10 seconds of data at about 20 Hz frame rate where the images were individually recorded as high quality JPEG images. This recording did contain some spatial compression, but no temporal compression which was the focus of the experiment. The collections were intentionally configured to have a relatively large amplitude rPPG signal by having a subject's finger back-lit by a phone's light. The same methodology executed on the synthetic data at varying bit rates and compression color spaces was evaluated on the sensor data. The results aligned with expectations based on the prior analysis and the HR signal was preserved using the RGB24 color space at lower bit rates compared to the YUV420 color space configuration (Figure 2.10).

This study demonstrated the potential value of selecting a compression color space that differs from the default setting when the application is known to be rPPG. The work was also motivated by the importance of amplitude preservation for remote pulse oximetry that builds beyond HR estimation. The intent of this work was not to report the results of all color space configuration options available for video compression as it relates to rPPG and SpO_2 estimation.

4.2   COTS Multi-Spectral Sensor for Physiological Monitoring

Some research related to physiological monitoring, such as the video compression study, pushes towards developing capabilities for consumer cameras with widespread use. Another valuable direction of work builds toward the usage of higher quality sensors that may be less likely for a consumer to own, but could be valuable in other settings such as a hospital. In a medical setting, the higher cost is more acceptable for more accurate physiological measurements. This second direction of work is what the experimentation and characterization of the Sentera 6x sensor focuses on.

These experiments served the purpose of path-finding and risk reduction for future research using this sensor by evaluating and characterizing items such as the feasibility of 5 Hz framerate, characteristics of exposure to inform experimental setups, and multi-spectral frame registration demonstration. The full list of observations and identified impacts to future studies are summarized in Table 3.3. Some key findings include inputs to future work such as avoiding exercise use cases due to aliasing, appropriately configuring the background of data collections to avoid skin saturation, and appropriate normalization of digital counts with dynamic exposure settings.

4.3 Opportunities for Future Work

Regarding video compression related to remote physiological monitoring, there are multiple pathways for future work. First, the characterization of a larger variety of color spaces and other video compression options could be executed and the experimental setup could be modified for a more common reflective rPPG setup with ambient lighting. As previously mentioned, the relatively high amplitude rPPG signal generated from the backlit finger was intended to simplify the concept demonstration. Future work using ambient light could extend the concept to an environment more comparable to state-of-the-art applications of rPPG.

Another direction for future work would collect uncompressed data in a configuration amenable to blood oxygen saturation estimation. While the importance of amplitude for blood oxygen saturation was a motivating factor for the video compression study, the amplitude preservation was directly measured and the impact on blood oxygen saturation estimation was approximated using principles of the RoR technique. Building off of this initial study, directly measuring the impact of video compression and the selection of color spaces on SpO_2 estimation would be a valuable direction for future work.

In regards to the Sentera 6x sensor, the most attractive areas of future work are related to remote pulse oximetry due to the additional and more precise spectral information available from the sensor. As identified through this work, there will be attributes of the sensor that must be appropriately addressed, but the multi-spectral property of the sensor still provides opportunities for future studies.

Going beyond the two focus areas of this thesis, there are other attributes of how data is stored in consumer cameras that could be better accounted for to improve the accuracy of remote

physiological monitoring, particularly in the case of remote pulse oximetry. The video compression work discussed the sub-sampling of the Cb and Cr color channels that is motivated by what the human visual system is sensitive to. There are other attributes of data storage such as gamma correction that could be properly inverted to improve accuracy. There has also been work related to the calibration of cameras to accurately reconstruct the RAW image that is much more closely related to the irradiance incident on the camera aperture compared to the values stored in an RGB file [35]. This procedure requires a relatively extensive calibration procedure that may only be reasonable to execute in a clinical setting as opposed to a general consumer. If this procedure was evaluated on collections and remote pulse oximetry algorithms were compared on the RGB and RAW data, it may provide a better characterization of the limits of what can be achieved with consumer cameras.

Bibliography

[1] Yitzhak Mendelson. Pulse oximetry: theory and applications for noninvasive monitoring. *Clinical chemistry*, 38(9):1601—1607, Sept. 1992.

[2] 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.

[3] Dangdang Shao, Chenbin Liu, Francis Tsow, Yuting Yang, Zijian Du, Rafael Iriya, Hui Yu, and Nongjian Tao. Noncontact monitoring of blood oxygen saturation using camera and dual-wavelength imaging system. *IEEE Transactions on Biomedical Engineering*, 63(6):1091–1098, 2016.

[4] John G. Webster. *Design of Pulse Oximeters*. CRC Press, Boca Raton, 1997.

[5] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, New York, 2022.

[6] Kenneth J. Arrow. Industrial cameras: Spectral sensitivity. Report, The Imaging Source, 2017.

[7] Wim Verkruysse, Marek Bartula, Erik Bresch, Mukul Rocque, Mohammed Meftah, and Ihor Kirenko. Calibration of contactless pulse oximetry. *Anesthesia & Analgesia*, 124:1, May 2016.

[8] Andreia Moço and Wim Verkruysse. Pulse oximetry based on photoplethysmography imaging with red and green light: Calibratability and challenges. *Journal of Clinical Monitoring and Computing*, 35:1–11, Feb. 2021.

[9] Christopher G. Scully, Jinseok Lee, Joseph Meyer, Alexander M. Gorbach, Domhnull Granquist-Fraser, Yitzhak Mendelson, and Ki H. Chon. Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering*, 59(2):303–306, 2012.

[10] Xinyi Ding, Damoun Nassehi, and Eric C. Larson. Measuring oxygen saturation with smartphone cameras using convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, 23(6):2603–2610, 2019.

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

[12] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam — a non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, 2011.

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

[14] Oliver T. Unke and Markus Meuwly. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of Chemical Theory and Computation*, 15(6):3678–3693, May 2019.

[15] Weixuan Chen and Daniel McDuff. DeepPhys: Video-based physiological measurement using convolutional attention networks, 2018.

[16] Daniel J. McDuff, Ethan B. Blackford, and Justin R. Estepp. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 63–70, 2017.

[17] Changchen Zhao, Chun-Liang Lin, Weihai Chen, and Zhengguo Li. A novel framework for remote photoplethysmography pulse extraction on compressed videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1380–138009, 2018.

[18] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. Systematic analysis of video-based pulse measurement from compressed videos. *Biomed. Opt. Express*, 12(1):494–508, Jan. 2021.

[19] Samruddhi Y. Kahu, Rajesh B. Raut, and Kishor M. Bhurchandi. Review and evaluation of color spaces for image/video compression. *Color Research & Application*, 44(1):8–33, 2019.

[20] Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. Standard, International Telecommunications Union, Geneva, CH, Mar. 2011.

[21] Parameter values for the HDTV standards for production and international programme exchange. Standard, International Telecommunications Union, Geneva, CH, June 2015.

[22] Parameter values for ultra-high definition television systems for production and international programme exchange. Standard, International Telecommunications Union, Geneva, CH, Oct. 2015.

[23] Jesse D. Kornblum. Using JPEG quantization tables to identify imagery processed by software. *Digital Investigation*, 5:S21–S25, 2008. The Proceedings of the Eighth Annual DFRWS Conference.

[24] Suramya Tomar. Converting video formats with FFmpeg. *Linux Journal*, 2006(146):10, 2006.

[25] Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. Standard, International Telecommunication Union, Geneva, Mar. 2011.

[26] Xin Tian, Chau-Wai Wong, Sushant M. Ranadive, and Min Wu. A multi-channel ratio-of-ratios method for noncontact hand video based SpO_2 monitoring using smartphone cameras. *IEEE Journal of Selected Topics in Signal Processing*, 16(2):197–207, Feb. 2022.

[27] Medical electrical equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment. Standard, International Organization for Standardization, Geneva, CH, Dec. 2017.

[28] Ary L. Goldberger, Luis A. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen C. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung K. Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):E215–20, Jun. 2000.

[29] Qiang Zhu, Mingliang Chen, Chau-Wai Wong, and Min Wu. Adaptive multi-trace carving for robust frequency tracking in forensic applications. *IEEE Transactions on Information Forensics and Security*, 16:1174–1189, 2021.

[30] Qiang Zhu, Chau-Wai Wong, Chang-Hong Fu, and Min Wu. Fitness heart rate measurement using face videos. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2000–2004, 2017.

[31] Max W. Hauser. Principles of oversampling A/D conversion. *J. Audio Eng. Soc*, 39(1/2):3–26, 1991.

[32] Davis E. King. Dlib-ml: A machine learning toolkit. 10:1755–1758, Dec. 2009.

[33] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*, pages 2564–2571, 2011.

[34] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, Jun. 1981.

[35] Seon J. Kim, Hai T. Lin, Zheng Lu, Sabine Süsstrunk, Stephen Lin, and Michael S. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2289–2302, 2012.