# Analyzing Pulse From Head Motions in Video

by

Guha Balakrishnan

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2014





Author .....
Department of Electrical Engineering and Computer Science
October 1, 2013

Certified by .....
John Guttag
Professor, Electrical Engineering and Computer Science
Thesis Supervisor

Certified by .....
Fredo Durand
Professor, Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Chair, Department Committee on Graduate Theses

# Analyzing Pulse From Head Motions in Video

by

Guha Balakrishnan

Submitted to the Department of Electrical Engineering and Computer Science
on October 1, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Engineering

## Abstract

We extract heart rate and beat lengths from videos by measuring subtle head oscillations that accompany the cardiac cycle. Our method tracks features on the head, temporally filters their trajectories and performs principal component analysis (PCA) to decompose the trajectories into a set of 1D component motions. It then chooses the component that is most periodic as the pulse signal. Finally, we identify peaks of the chosen signal, which correspond to heartbeats. When evaluated on 18 subjects our approach reported heart rates nearly identical to an electrocardiogram (ECG) device for all subjects and obtained similar beat length distributions for 17. In addition we obtained pulse rate from videos of the back of the head and of sleeping newborns.

Initial findings also show that our method can measure heart rate from body parts other than the head and can measure respiration rate by selecting a different frequency band. Finally, we present visualization techniques such as motion magnification for subjective analysis of body motions.

Thesis Supervisor: John Guttag
Title: Professor, Electrical Engineering and Computer Science

Thesis Supervisor: Fredo Durand
Title: Professor, Electrical Engineering and Computer Science

# Acknowledgments

I first thank my advisors John Guttag and Fredo Durand for making this research a fun endeavor and bringing their knowledge, guidance and optimism to the work.

I thank the members of the Data Driven Medical Group: Joel, Anima, Garthee, Jenna, Jen, Amy and Yun for their advice and comments over the past couple of years. They also patiently sat through my video recording sessions, which were critical for my experiments. Thanks to Collin Stultz who helped keep our work on track from the medical perspective. I also appreciate the feedback from Bill Freeman, Tiam Jaroensri, Michael Rubinstein and Neal Wadhwa. Their related research in the computer vision/medical space was a motivating force for me.

Finally I am grateful to my family, who have always taken an interest in my academic progress. They made it very easy for me to concentrate on my graduate work and supported me in many ways.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Heart rate is a critical vital sign for medical diagnosis. There is growing interest in extracting it without contact, particularly for populations such as premature neonates and the elderly for whom the skin is fragile and damageable by traditional sensors. Furthermore, as the population ages, continuous or at least frequent monitoring outside of clinical environments can provide doctors with not just timely samples but also long-term trends and statistical analyses. Acceptance of such monitoring depends in part on the monitors being non-invasive and non-obtrusive.

In this thesis, we exploit subtle head oscillations that accompany the cardiac cycle to extract information about cardiac activity from video recordings. In addition to providing an unobtrusive way of measuring heart rate, the method can be used to extract other clinically useful information about cardiac activity, such as the subtle changes in the length of heartbeats that are associated with the health of the autonomic nervous system. Our method works with typical video recordings and is not restricted to any particular view of the head.

The cyclical movement of blood from the heart to the head via the abdominal aorta and the carotid arteries causes the head to move in a periodic motion. Our algorithm detects pulse from this movement. Our basic approach is to track feature points on a person's head, filter their positions by a temporal frequency band of interest, and use principal component analysis (PCA) to find a periodic signal caused by pulse. We extract an average pulse rate from this signal by examining its frequency

spectrum and obtain precise beat locations with a simple peak detection algorithm.

## 1.1 Related Work

Our method is an alternative to the extraction of pulse rate from video via analysis of the subtle color changes in the skin caused by blood circulation [18, 26]. These methods average pixel values for all channels in the facial region and temporally filter the signals to an appropriate band. They then either use these signals directly for analysis [26] or perform ICA to extract a single pulse wave [18]. They find the frequency of maximal power in the frequency spectrum to provide a pulse estimate. Philips produced a commercial app that detects pulse from color changes in real-time [17]. These color-based detection schemes require that facial skin be exposed to the camera. In contrast our approach is not restricted to a particular view of the head, and is effective even when skin is not visible.

There have also been studies on non-invasive pulse estimation using modalities other than video such as thermal imagery [6], photoplethysmography (measurement of the variations in transmitted or reflected light in the skin) [29] and laser and microwave Doppler [25, 7]. Noncontact assessment of heart rate variability (HRV), an index of cardiac autonomic activity, presents a greater challenge and few attempts have been made [21, 13, 15]. A common drawback of these non-invasive systems is that they are expensive and require specialized hardware.

The analysis of body motion in videos has been used in different medical contexts, such as the measurement of respiration rate from chest movement [23, 17], or the monitoring of sleep apnea by recognizing abnormal respiration patterns [28]. Motion studies for diseases include identification of gait patterns of patients with Parkinson's disease [4], detection of seizures for patients with epilepsy [16] and early prediction of cerebral palsy [1]. The movements involved in these approaches are larger in amplitude than the involuntary head movements due to pulse.

Our work is inspired by the amplification of imperceptible motions in video [30, 12]. But whereas these methods make small motions visible, we want to extract

quantitative information about heartbeats.

The idea of exploiting Newton's Third Law to measure cardiac activity dates back to at least the 1930's, when the ballistocardiogram (BCG) was invented [20]. The subject was placed on a low-friction platform, and the displacement of the platform was used to measure cardiac output. The BCG was never widely used in clinical settings. Other clinical methods using a pneumatic chair and strain-sensing foot scale have also been successful under laboratory conditions[11, 10]. Ballistocardiographic head movement of the sort studied here has generally gained less attention. Such movement has been reported during studies of vestibular activity and as an unwanted artifact during MRI studies [2]. Recently, He *et al.*[8] proposed exploiting head motion measured by accelerometers for heart rate monitoring as a proxy for traditional BCG.

## 1.2 Contributions

This thesis makes the following contributions:

1. Development of a method for extracting pulse rate and heartbeat variability from videos by analyzing ballistocardiac motions of the head. Our method tracks feature points on the head, filters their trajectories to a frequency band of interest, and uses PCA to decompose the trajectories into 1-D signals. The most periodic 1-D signal is then chosen to calculate pulse rate and find beat intervals. Our method is not restricted to any view of the face, and can even work when the face is occluded.

2. Validation of our method on video recordings

   (a) Our system returned pulse rates nearly identical to an electrocardiogram (ECG) device for 18 subjects.

   (b) We are the first to evaluate beat-length distributions from video as a way to evaluate variability. We captured very similar beat length distributions to the ECG for our subjects.

(c) We extracted pulse rate from videos of the back of the head and when a subject was wearing a mask.

(d) We extracted pulse rate for several videos of sleeping infants.

3. Noise analysis of our method and a comparison to a color-based pulse detection scheme [18].

4. Results showing that our method can also be used to measure respiration rate and extract pulse from movements in the chest and carotid artery. While our techniques were developed for measuring pulse from head motions, we tried to keep our technique general enough to not be limited to the head alone.

5. Presentation of visualization techniques such as frequency trajectory plots and motion magnification for more subjective evaluation of body motions. These methods are useful for exploratory analysis of motions in video.

## 1.3 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 provides background of relevant physiology and computer vision techniques. Chapter 3 presents our method. In Chapter 4 we evaluate our method on video recordings. In Chapter 5 we present initial results showing how our method can be used for other applications. In Chapter 6 we present visualization techniques for subjective evaluation of body motions. Finally, we provide a summary and conclude in Chapter 7.

# Chapter 2

# Background

This section provides a brief background on several key topics related to our study. First we present the main elements of the cardiac cycle and the ballistocardiac forces that cause body and head motions. We then discuss the clinical significance of heart beat variability. We then give a simple overview of feature tracking, an important element of our pulse detection system. Finally we summarize a recent method on pulse extraction from changes in skin color to which we compare our work in our experimental evaluation.

## 2.1 The Cardiac Cycle and Ballistocardiac Forces

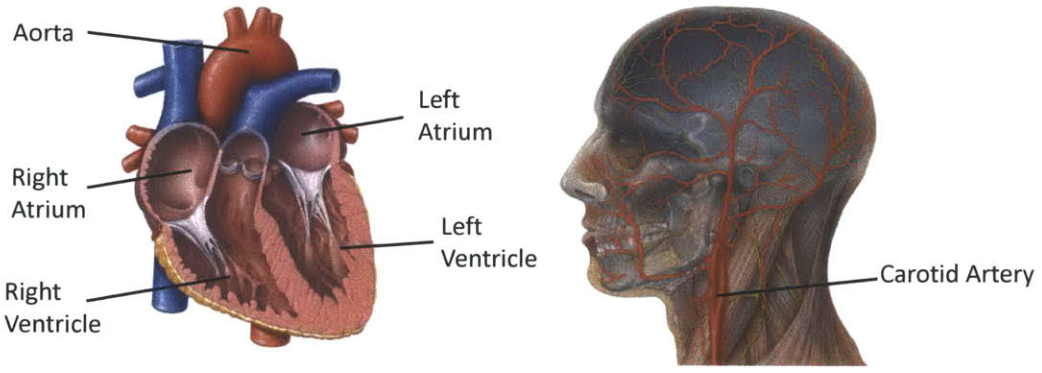



Figure 2-1: Blood flows from the heart to the head via the carotid arteries on either side of the head [14].

The heart is is a hollow muscle that pumps blood throughout the body by repeated, cyclical contractions. It is composed of four chambers: the left/right ventricle and the left/right atrium (Fig. 2-1). During the phase of the cardiac cycle known as *diastole*, the ventricles relax and allow blood to flow into them from the atria. In the next phase known as *systole*, the ventricles contract and pump blood to the pulmonary artery and aorta. The aorta in turn transports blood to the rest of the body. The head and neck receive blood from the aorta via the common carotid arteries, which further divide into the internal and external carotid arteries in the neck.

We are not certain how much of head motion we detect is attributable to the large acceleration of blood in the aorta compared to the localized acceleration of blood in the carotid arteries moving into the head. He *et al.* [8] measured ballistocardiac head motions with an accelerometer-based device to be on the order of $10\,\text{mG}(0.098\,\frac{\text{m}}{\text{s}^2})$, but it is unclear what fraction of this movement is attributed to aortic and carotid blood flow forces. To understand this better, we will present derivations of the magnitude of each of these forces in Sections 2.1.1 and 2.1.2. The calculations are simplified, ignoring many details about fluid dynamics and physiology. From a biomechanical standpoint, the head-neck system and the trunk can be considered as a sequence of stacked inverted pendulums. This structure allows the head unconstrained movement in most axes, making calculations of the system's motion complicated.

### 2.1.1 Aortic Force

During systole a healthy adult aorta is capable of impelling a volume of $70 - 80\,\text{ml}$ $(0.07 - 0.08\,\text{kg})$ of blood at considerable speed $(1\,\frac{\text{m}}{\text{s}})$ and acceleration $(20\,\frac{\text{m}}{\text{s}^2})$ [19]. An average adult male weighs $70\,\text{kg}$. Using Newton's $2^{nd}$ and $3^{rd}$ laws we derive the approximate acceleration of the body due to the aortic forces ($a_{body}$):

$$|\mathbf{F_{body}}| = |\mathbf{F_{blood}}|$$
$$|\mathbf{a_{body}} * m_{body}| = |\mathbf{a_{blood}} * m_{blood}|$$
$$|\mathbf{a_{body}} * 70\,\mathrm{kg}| = 20\,\frac{\mathrm{m}}{\mathrm{s}^2} \cdot 0.07\,\mathrm{kg}$$
$$|\mathbf{a_{body}}| = 0.02\,\frac{\mathrm{m}}{\mathrm{s}^2}$$

Thus the approximate acceleration of the body due to the aortic forces is $0.02\,\frac{\mathrm{m}}{\mathrm{s}^2}$. Ballistocardiography produces a graphical representation of body movement due to this force for a subject lying down on a low-friction platform (Fig. 2-2).

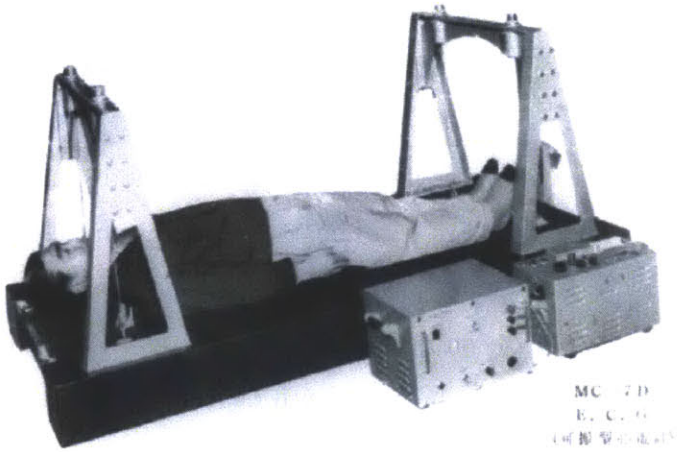


Figure 2-2: An old model of a ballistocardiogram where a subject lies on a low-friction platform. The displacement and/or acceleration of the body is measured to infer cardiac output. Image courtesy of Nihon Kohden.

### 2.1.2 Carotid Force

In addition to aortic blood acceleration, there is the smaller force resulting from blood flowing through the carotid arteries into the head. Although there are studies measuring blood flow, velocity and acceleration of blood in the carotid, we have not found similar experimental measurements on the force imparted by carotid blood flow

on the head. According to one study, blood velocity in the common carotid increases from approximately $0.02\,\frac{\mathrm{m}}{\mathrm{s}}$ to $0.11\,\frac{\mathrm{m}}{\mathrm{s}}$ in 0.1 s, or an acceleration of $0.9\,\frac{\mathrm{m}}{\mathrm{s}^2}$ [9]. The blood mass transported in this period is roughly 13 ml or 0.013 kg. Assuming both carotid arteries are identical and that the head is 5 kg, we use a similar derivation to the one used in Section 2.1.1 to find that the head should accelerate roughly $0.005\,\frac{\mathrm{m}}{\mathrm{s}^2}$ assuming independence from the rest of the body.

### 2.1.3 Heart Rate and Heart Rate Variability

Pulse rate captures the average number of cardiac cycles over a period of time (e.g., 30 seconds). It is useful primarily for detecting acute problems. There is a growing body of evidence [22] that measuring beat-to-beat variations provides additional information with long-term prognostic value. The most established of these measures is heart rate variability (HRV). HRV measures the variation in the length of individual normal (sinus) heartbeats. It provides an indication of the degree to which the sympathetic and parasymathetic nervous systems are modulating cardiac activity. To measure HRV, the interarrival times of beats must be accurately measured, which can be determined by locating the "R" peaks in successive beats in an ECG. A lack of sufficient variation when the subject is at rest suggests that the nervous system may not perform well under stress. Patients with decreased HRV are at an increased risk of adverse outcomes such as fatal arrhythmias.

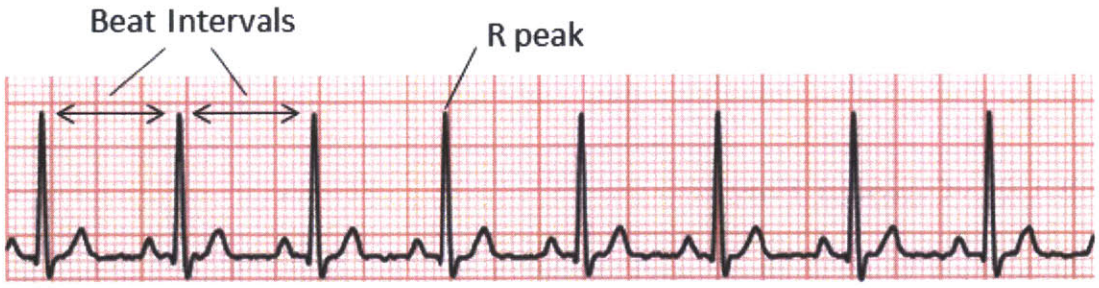



Figure 2-3: Example of typical ECG signal. Differences between R peaks in successive beats of an ECG can be used to calculate heart rate variability or HRV.

## 2.2 Optical Flow and the Lucas-Kanade Algorithm

Our approach relies on optical flow tracking to quantify the motion of a head in a video. Optical flow is the pattern of motion of objects, surfaces and edges in a visual scene. Optical flow methods try to calculate the motion of pixel locations between two image frames that are taken at times $t$ and $t + \Delta t$. These methods are based on local Taylor series approximations of the image signal. For a $2D + t$ dimensional case a pixel at location $(x, y, t)$ with intensity $I(x, y, t)$ will have moved by $\Delta x$, $\Delta y$ and $\Delta t$ between the two image frames and the following image constraint equation can be given:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, x + \Delta z) \tag{2.1}$$

Assuming the movement to be small (subpixel or a few pixels in amplitude), the constraint can be rewritten using a Taylor series:

$$I(x + \Delta x, y + \Delta y, x + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \tag{2.2}$$

From these equations it follows that:

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0 \tag{2.3}$$

or

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0 \tag{2.4}$$

This can be rewritten to form the optical flow equation:

$$I_x V_x + I_y V_y = -I_t \tag{2.5}$$

where $V_x$, $V_y$ are the $x$ and $y$ components of the velocity or optical flow of $I(x, y, t)$ and $I_x$, $I_y$ and $I_t$ are the partial derivatives of the image at $(x, y, t)$.

The Lucas-Kanade method of optical flow assumes the velocity is approximately

constant within a small neighborhood of the point $p$ under consideration. The optical flow equation is made to hold for all points within the neighborhood and solutions to $V_x$ and $V_y$ are obtained using a least squares fit. In our work we use the Lucas-Kanade implementation provided in OpenCV [3]. This implementation is a sparse feature point tracker, meaning that it finds the motions of selected point locations in the image rather than finding a dense motion field for the entire image. The sparse tracker worked well for our application and is also much quicker than a dense optical flow algorithm.

For scenes with large motions, the Taylor approximation in Eq. 2.2 does not hold. In these cases one can use a coarse-to-fine optical flow scheme with image pyramids consisting of downsampled versions of the original images. The feature point velocities are refined from the coarsest image down to the finest (original) image. Since the motions we study are small, we did not use pyramids for our work.

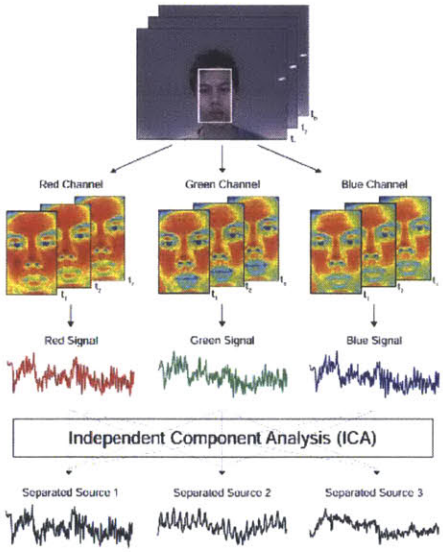



Figure 2-4: Overview of color-based pulse detection method (image taken from [18]). A face is found from the first video frame to construct a region of interest (ROI). The red, green and blue channels are spatially averaged in the ROI in each frame to form three signals. ICA is applied on these signals to obtain three independent source signals. The most periodic source signal is chosen for analysis.

## 2.3 Color-Based Pulse Detection from Video

Blood circulation causes volumetric changes in blood vessels that modify the path length of ambient light. This is the basic premise of plethysmography. The red, green, and blue (RGB) sensors of a video camera can pick up this plethysmographic signal mixed with other fluctuations of light caused by artifacts. In addition, because hemoglobin absorptivity varies across the visible spectral range, each of these sensors records a mixture of these sources with different weights. The task of pulse signal extraction is to reconstruct the signal of interest from these channels. A work that has gained recent interest separates the plethysmographic signal from the noise using independent component analysis (ICA)[18]. The method (see Fig. 2-4) spatially averages the R, G, and B values from the facial area in each video frame to form three signals. Then, ICA is used to decompose the signals into 3 independent source signals. The source with the largest peak in the power spectrum is chosen as the pulse signal. Finally, the signal is smoothed and interpolated to 256 Hz for HRV analysis. When they evaluated their method on 12 subjects, they produced accurate heart rate and HRV measures.

Analysis of HRV was performed by power spectral density (PSD) estimation using the Lomb periodogram. They found the low frequency (LF) and high frequency (HF) powers which reflect different properties of the sympathetic and parasympathetic influences on the heart. The video recordings used in their analysis were roughly a minute in length, similar to the length of recordings in our work. However, this is a much shorter duration than the hours of ECG recordings used to calculate HRV measures in practice and it is not clear whether it is clinically meaningful. Therefore, in our work, we directly evaluate heartbeat length distributions instead of using the standard HRV measures.

# Chapter 3

# Method

Our method takes an input video of a person's head and returns a pulse rate as well as a series of beat locations that can be used for the analysis of beat-to-beat variability. We first extract the motion of the head using feature tracking. We then isolate the motion corresponding to the pulse and project it onto a 1D signal that allows us to extract individual beat boundaries from the peaks of the trajectory. For this, we use PCA and select the component with the most periodic 1D projection. Finally we extract the beat locations as local extrema of the chosen 1D signal.

Fig. 3-1 presents an overview of the technique. We assume the recorded subject is stationary and sitting upright for the duration of the video. We start by locating the head region and modeling head motion using trajectories of tracked feature points. We use the vertical component of the trajectories for analysis. The trajectories have extraneous motions at frequencies outside the range of possible pulse rates, and so we temporally filter them. We then use PCA to decompose the trajectories into a set of independent source signals that describe the main elements of the head motion. To choose the correct source for analysis and computation of the duration of individual beats, we examine the frequency spectra and select the source with the clearest main frequency. We use this criterion because pulse motion is the most periodic motion of the head within the selected frequency band. Average pulse rate is identified using this frequency. For more fine-grained analysis and calculation of beat durations, we perform peak detection in the time-domain.

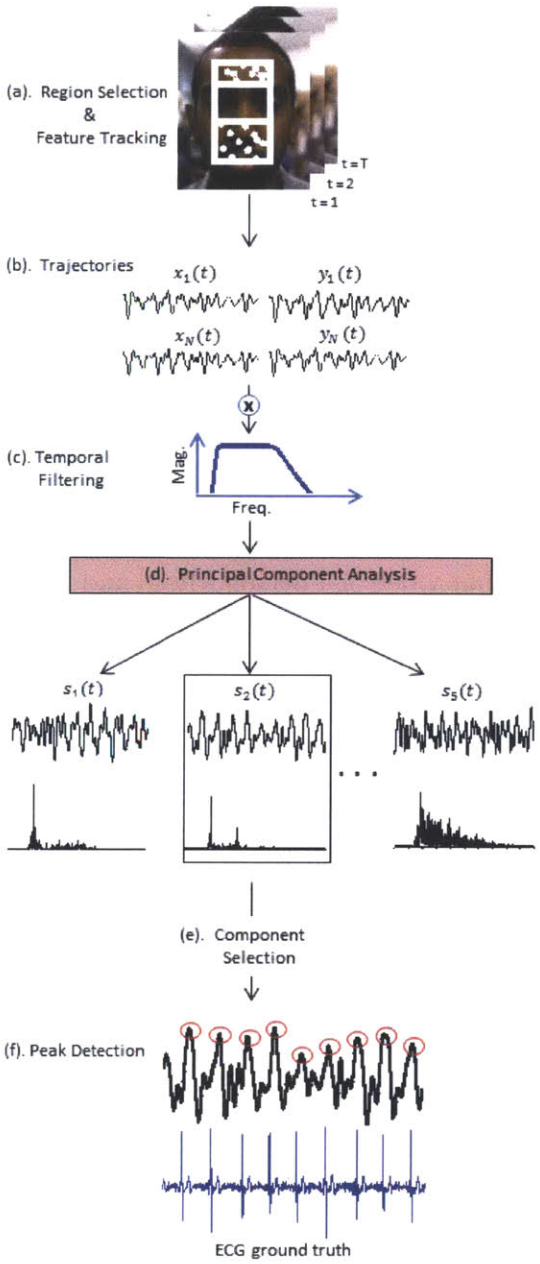



Figure 3-1: Overview of our pulse estimation approach. (a) A region is selected within the head and feature points are tracked for all frames of the video. (b) The horizontal and vertical components are extracted from each feature point trajectory. (c) Each component is then temporally filtered to remove extraneous frequencies. (d) PCA decomposes the trajectories into a set of source signals $s_1, s_2, s_3, s_4, s_5$. (e) The component that has clearest main frequency is selected. (f). Peak detection identifies the beats of the signal.

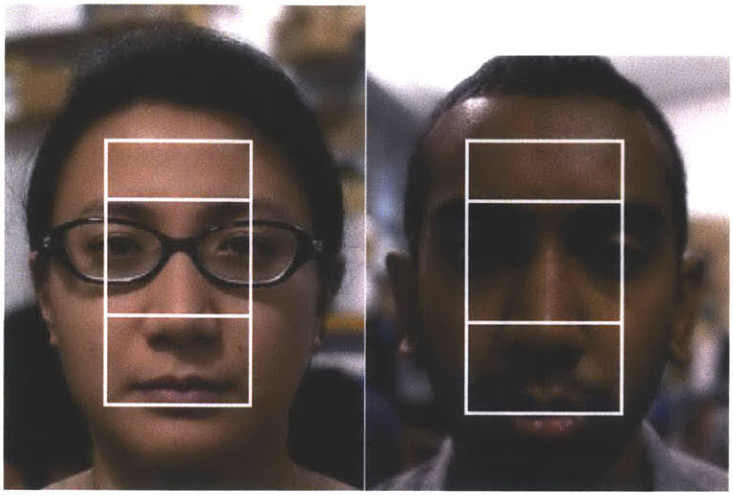


Figure 3-2: The region of interest for subjects 1 and 2. The region encompasses the middle of the face.

## 3.1 Region Selection

We find a region of interest containing the head and track feature points within the region. For videos where the front of the face is visible, we use the Viola Jones face detector [27] from OpenCV 2.4 [3] to first find a rectangle containing the face. We use the middle 50% of the rectangle widthwise and 90% heightwise from top in order to ensure the entire rectangle is within the facial region. We also remove the eyes from the region so that artifacts caused by blinking do not affect our results. To do this we found that removing the subrectangle spanning 20% to 55% heightwise works well (see Fig. 3-2). For videos where the face is not visible, we mark the region manually.

## 3.2 Feature Point Selection

We measure the movement of the head throughout the video by selecting and tracking feature points within the region. We use the OpenCV Harris corner detector (the *goodFeaturesToTrack* function) to select the points. We set the parameters of this function to ensure that we select 1500 points evenly distributed around the face.

## 3.3 Feature Point Tracking

We apply the OpenCV Lucas Kanade tracker between frame 1 and each frame $t = 2 \cdots T$ to obtain the location time-series $\langle x_n(t), y_n(t) \rangle$ for each point $n$. We used a window size of 40 pixels around each feature point for tracking. This window size is $\frac{1}{125}$ the area of the face rectangle found by the detector for our subjects on average. It is large enough to capture local features around the nose/mouth area.

Many of the feature points can be unstable and have erratic trajectories. To retain the most stable features we find the maximum distance traveled by each point between consecutive frames and discard points with a distance exceeding the $75^{th}$ percentile.

## 3.4 Temporal Filtering

Not all frequencies of the trajectories are useful for pulse detection. A normal adult's resting pulse rate falls within $[0.75, 2]$ Hz, or $[45, 120]$ beats/min. We found that frequencies lower than 0.75 Hz negatively affect our system's performance. This is because low-frequency movements like respiration and changes in posture have high amplitude and dominate the trajectories of the feature points. However, harmonics and other frequencies higher than 2 Hz provide useful precision needed for peak detection. Taking these elements into consideration, we filter each $x_n(t)$ and $y_n(t)$ to a passband of $[0.75, 5]$ Hz. For babies, who have faster pulses, we use a passband of $[1.25, 5]$ Hz. We use a $5^{th}$ order butterworth filter for its maximally flat passband.

## 3.5 PCA Decomposition

The underlying source signal of interest is the movement of the head caused by the cardiovascular pulse. The feature point trajectories are a mixture of this movement as well as other motions caused by sources like respiration, vestibular(balancing) activity and changes in facial expression. Each subject exhibits different motions. Fig. 3-3 shows examples of the total energy of the feature point trajectories at different

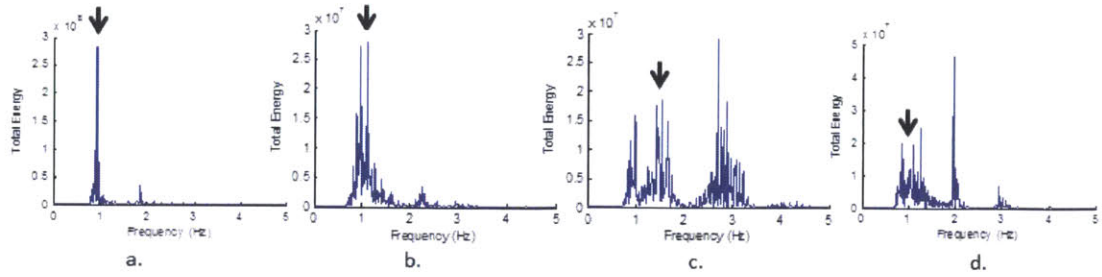



Figure 3-3: Examples of the combined energy spectra of feature points for 4 different subjects. (a) is an example where the pulse frequency is the only dominant frequency. (b) is an example where there is another frequency with nearly as much energy as pulse. (c) is an example where the harmonic of the pulse frequency has more energy than the pulse frequency. (d) is an example where the energy at the pulse frequency is smaller than several other peaks in the spectrum. We were able to get accurate pulse signals for all subjects using our method which separates the combined motion of the feature points into submotions using PCA.

frequencies for four different subjects. The black arrow indicates where the true pulse rate is. In the first case (a), the dominant frequency corresponds to pulse. This is the easiest case. In the second case (b) the pulse still has maximum energy but there is another frequency with nearly as much energy. In the third case (c), the harmonic of the pulse has greater energy than the pulse itself. In case (d) the pulse is completely masked by extraneous motions. The task of developing a method to extract a reliable pulse for these different cases is challenging.

Our task is to decompose the mixed motions of the feature points into subsignals to isolate pulse. To do this we consider the multidimensional position of the head at each frame as a separate data point and use PCA to find a set of dimensions along which the position varies. We then select a dimension on which to project the position time-series to obtain the pulse signal.

Formally, given $N$ feature points, we represent the $N$-dimensional position of the head at frame $t$ as $m_t = [x_1(t), x_2(t), \cdots, x_N(t), y_1(t), y_2(t), \cdots, y_N(t)]$. The mean and the covariance matrix of the positions are defined by:

$$\bar{m} = \frac{1}{|T|} \sum_{i=1}^{T} m_i \tag{3.1}$$

$$\Sigma_m = \frac{1}{T}\sum_{i=1}^{T}(m_i - \bar{m})(m_i - \bar{m})^T \tag{3.2}$$

PCA finds the principal axes of variation of the position as the eigenvectors of $\Sigma_m$:

$$\Sigma_m \Phi_m = \Phi_m \Lambda_m \tag{3.3}$$

where $\Lambda_m$ denotes a diagonal matrix of the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_N$ corresponding to the eigenvectors in the columns of $\Phi_m$, $\phi_1, \phi_2, \cdots, \phi_N$.

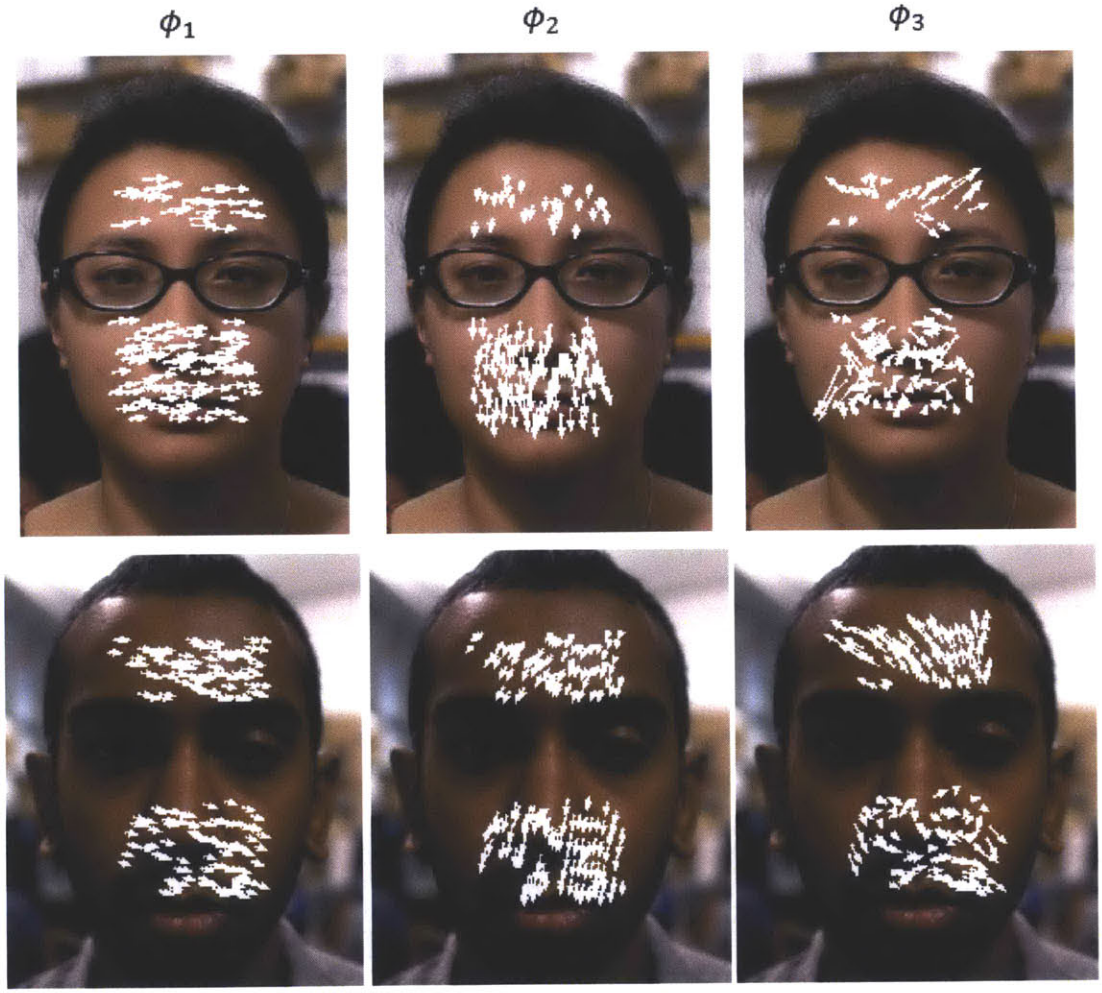



Figure 3-4: Examples of the first three eigenvectors for two subjects. Each white arrow on a face represents the magnitude and direction of a feature point's contribution to that eigenvector. The eigenvector decomposition is unique to each video.

Fig. 3-4 displays the first three eigenvectors for two of the subjects. Each eigenvector represents the 2N-dimensional direction and magnitude of movement for the feature points. The eigenvectors differ for each video. We obtain the 1-D position signal $s_i(t)$ by projecting the position time-series onto $\phi_i$:

$$s_i(t) = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_T \end{pmatrix} \cdot \phi_i \tag{3.4}$$

There are periods in the video during which the head moves abnormally (e.g. swallowing, adjustments in posture). Such movement adds variance to the position vectors, thereby affecting the PCA decomposition. To deal with this we discard a percentage $\alpha$ of the $m_t$ with the largest $L2$-norms before performing PCA. However, all of the $m_t$ must still be used in the projection step (Eq. 3.4) to produce a complete signal. We set $\alpha$ at 15% for our experiments.

A popular alternative to PCA is independent component analysis (ICA), which decomposes signals based on higher order statistics. We did not see any improvement in our results when using ICA.

## 3.6 Signal Selection

The question remains of which eigenvector to use for pulse signal extraction. The eigenvectors are ordered such that $\phi_1$ explains the most variance in the data, $\phi_2$ explains the second most, and so on. Fig. 3-5 shows the percentage of total variance attributed to the first 5 eigenvectors averaged over all subjects. The first two eigenvectors account for 56% and 36% of total variance on average. This makes intuitive sense when looking at Fig. 3-4, which shows that the first two components tend to capture most of the horizontal and vertical motions of the head. However, this does not mean that $s_1$ or $s_2$ is the clearest signal for analysis. In fact, it is likely that these signals are not the best because they tend to capture extraneous head motions along

with pulse motion.

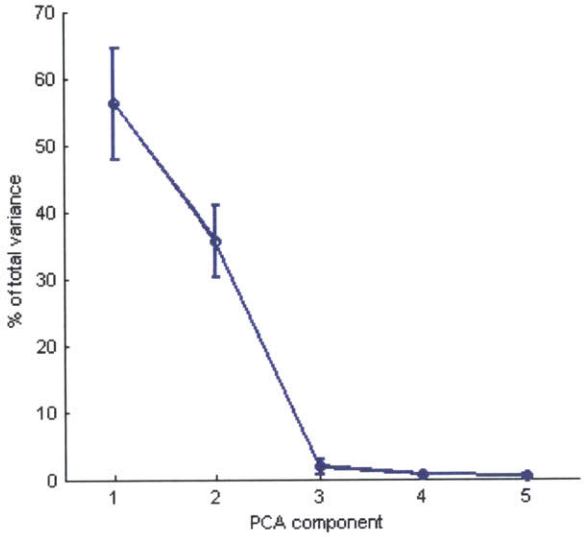



Figure 3-5: Percentage of total variance explained by the first five PCA components, averaged over all subjects in our experiments. Bars are standard deviations. The first component explains most of the variance for all subjects. However, the first component may not produce the best signal for analysis.

We instead choose the eigenvector $\phi_i$ with the signal $s_i$ that is most periodic. We define a signal's periodicity, $\rho$, as the percentage of total spectral energy accounted for by frequencies within 0.05 Hz of the dominant frequency and 0.05 Hz of its first harmonic. The energy spectrum of a signal is calculated using the Discrete Fourier Transform (DFT). If $S_i^j$ is the $j^{th}$ complex DFT coefficient of signal $s_i$, then the energy at $j$ is $(Re(S_i^j))^2 + (Im(S_i^j))^2$.

We found it unnecessary to consider any signals beyond the first five, i.e. $s_1, ..., s_5$ for any of our subjects. We label the maximum frequency of the chosen signal as $f_{pulse}$ and approximate the pulse rate as $60 * f_{pulse}$ beats per minute.

For the 18 subjects in our experiments the first five components were chosen the following number of times: 3, 1, 13, 1, and 0. The third component was chosen for 14 of the 18 videos. We hypothesize that this is because the first two components remove much of the extraneous motions as seen in Fig. 3-3 leaving a purer residual motion for the third component to capture.

## 3.7 Peak Detection

Pulse rate alone is not, of course, sufficient to fully evaluate the cardiovascular system. Clinicians often assess beat-to-beat variations to form a more complete picture. To allow for such analysis, we perform peak detection on the selected PCA component signal. Since a modern ECG device operates around 250 Hz to capture heart rate variability and our videos were only shot at 30 Hz, we first apply a cubic spline interpolation to the signal to increase its sampling rate to 250 Hz.

The peaks are close to $\frac{1}{f_{pulse}}$ seconds apart with some variability due to the natural variability of heartbeats, variations of the head motion, and noise. We label each sample in the signal as a peak if it is the largest value in a window centered at the sample. We set the length of the window (in samples) to be $round(\frac{250}{f_{pulse}})$. Fig. 3-6 shows examples of signals outputted by our method. Peaks are marked by circles.

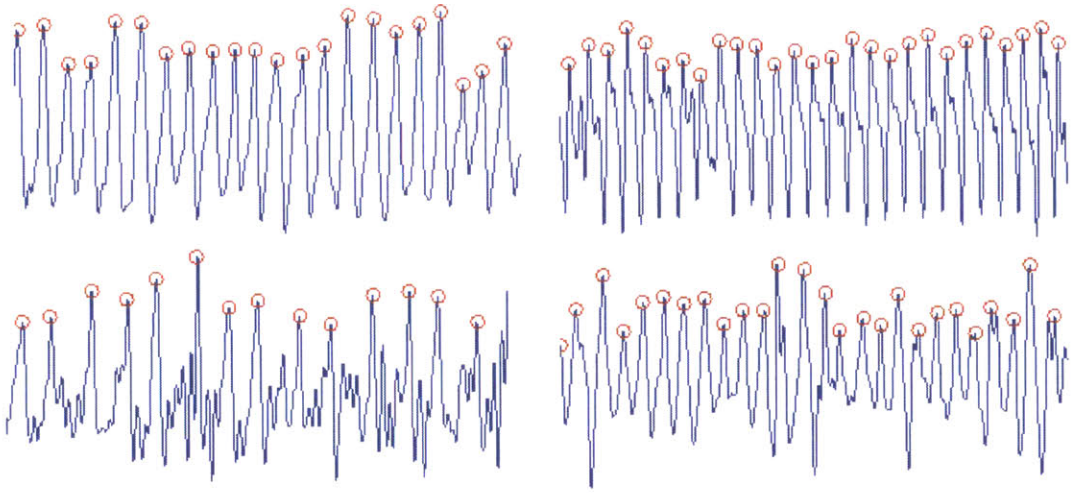


Figure 3-6: Examples of motion signals outputted by our method. Red circles are peaks.

# Chapter 4

# Experiments

We implemented our approach in MATLAB. Videos were shot with a Panasonic Lumix GF2 camera in indoor, unisolated environments with varying lighting conditions. All videos had a frame rate of 30 frames per second, 1280 x 720 pixel resolution and a duration of 70-90 seconds. Videos were saved in MJPEG format. During the recordings, subjects were told to sit still as possible while looking forward at the camera (for frontal videos) and away from the camera (side/back views). Fig. 4-1 shows frames from frontal videos of all 18 subjects. The subjects varied in gender (7 female, 11 male) and skin color. They ranged from 23-32 years of age and were all seemingly healthy. We connected subjects to a wearable ECG monitor [5] for ground truth comparison. This device has a sampling rate of 250 Hz and three electrodes that we placed on the forearms.

## 4.1 Visible Face

We extracted pulse signals from the 18 subjects with a frontal view of the face (Fig. 4-1). We calculate our program's average pulse rate using the frequency of maximal power for the selected PCA component. Similarly, we compute the true pulse rate by finding the main frequency of the ECG spectrum. Table 4.1 presents our results. The average rates are nearly identical to the true rates for all subjects, with a mean absolute difference of 1.4%. The number of peaks were also close to ground truth

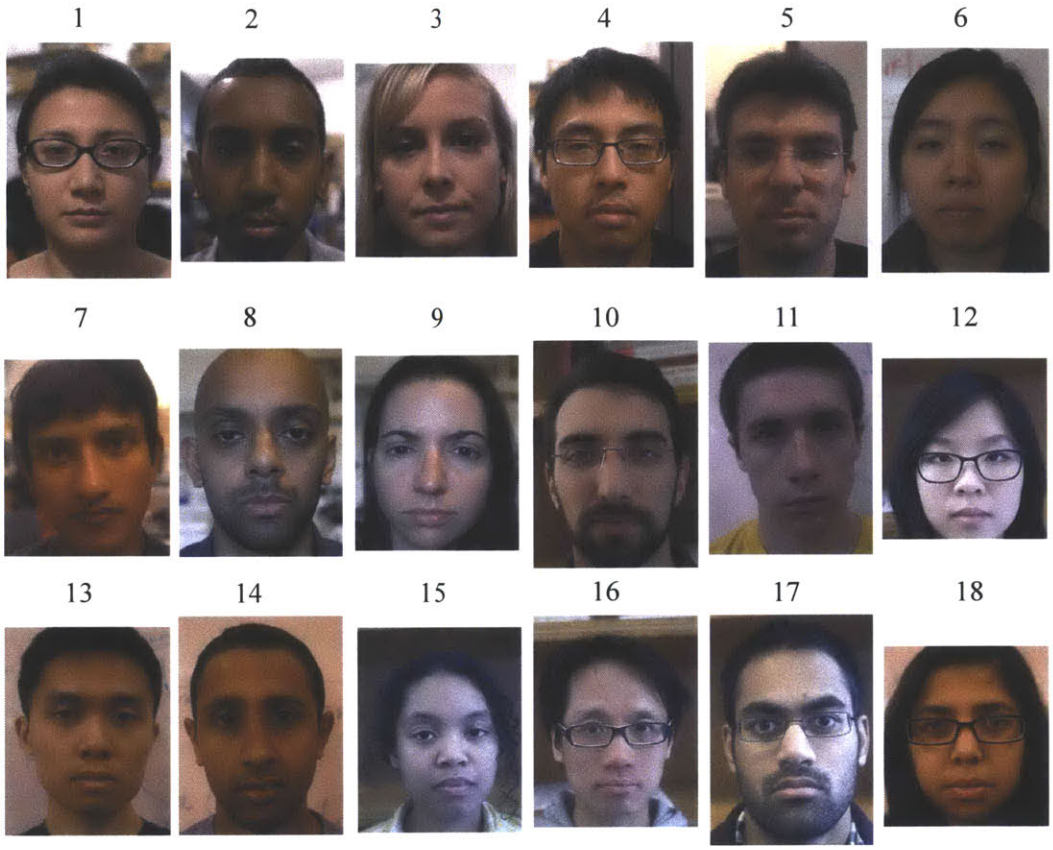



Figure 4-1: Reference frames from videos of all 18 subjects. Subjects varied in skin color, gender and age.

values, with a mean absolute difference of 2.1%.

We also evaluate the ability of our signal to capture subtle heart rate variability. Clinically meaningful HRV measures typically use 10-24 hours of ECG data. Therefore we did not attempt to compute any of these for our 90 second videos. Instead, we compare the distributions of time between successive peaks for each signal. Incorrect or missed peaks can introduce spurious intervals too large or small to be caused by the natural variations of the heart. We account for these cases by only considering intervals with a length within 25% of the average detected pulse period.

We use the Kolmogorov-Smirnov (KS) test to measure the similarity of the distributions, with the null hypothesis being that the observations are from the same distribution. Table 4.1 presents the results. At a 5% significance level, 17 of the 18 pairs of distributions were found to be similar. Fig. 4-2 presents histograms of 5 of

Table 4.1: Results when comparing the beat length distributions of the ECG and our method. Presented are the means ($\mu$) and standard deviations ($\sigma$) of the distributions along with $p$-value of the Kolmogorov-Smirnov test measuring distribution similarity. 17 of the 18 pairs of distributions were not found to be significantly different ($p >= 0.05$)

| Sub. | ECG | | | Motion | | | Similarity |
|---|---|---|---|---|---|---|---|
| | Pulse Rate | # of beats | $\mu(\sigma)$ | Pulse Rate (% err) | # of beats (% err) | $\mu(\sigma)$ | $p$-value (high is good) |
| 1 | 66.0 | 99 | 0.91(0.06) | 66.0(0) | 99(0) | 0.91(0.07) | 0.78 |
| 2 | 55.3 | 83 | 1.08(0.08) | 56.0(1.2) | 82(1.2) | 1.09(0.10) | 0.96 |
| 3 | 81.3 | 122 | 0.73(0.04) | 82.7(1.7) | 119(2.5) | 0.73(0.08) | 0.16 |
| 4 | 44.0 | 66 | 1.34(0.19) | 46.0(4.6) | 66(0) | 1.33(0.19) | 0.92 |
| 5 | 95.3 | 143 | 0.62(0.03) | 96.0(0.7) | 143(0) | 0.62(0.07) | <0.01 |
| 6 | 78.0 | 91 | 0.76(0.04) | 78.9(1.2) | 91(0) | 0.76(0.04) | 0.94 |
| 7 | 73.3 | 110 | 0.81(0.05) | 72.0(1.8) | 98(10.9) | 0.83(0.06) | 0.14 |
| 8 | 59.3 | 89 | 1.01(0.04) | 58.7(1.1) | 89(0) | 1.00(0.10) | 0.06 |
| 9 | 56.6 | 85 | 1.04(0.07) | 58.7(3.6) | 83(2.4) | 1.04(0.12) | 0.34 |
| 10 | 78.9 | 100 | 0.75(0.04) | 82.1(4.1) | 98(2.0) | 0.75(0.04) | 0.60 |
| 11 | 84.6 | 127 | 0.70(0.06) | 85.3(0.8) | 116(8.7) | 0.71(0.07) | 0.57 |
| 12 | 63.3 | 95 | 0.94(0.08) | 62.7(1.0) | 95(0) | 0.94(0.09) | 0.72 |
| 13 | 60.0 | 90 | 0.99(0.04) | 60.0(0) | 90(0) | 0.99(0.08) | 0.27 |
| 14 | 60.7 | 91 | 0.98(0.11) | 60.7(0) | 87(4.4) | 0.99(0.12) | 0.74 |
| 15 | 80.7 | 121 | 0.74(0.05) | 81.3(0.8) | 117(3.3) | 0.75(0.06) | 0.62 |
| 16 | 74.7 | 112 | 0.80(0.05) | 76.0(1.8) | 110(1.8) | 0.80(0.05) | 0.87 |
| 17 | 50.0 | 75 | 1.18(0.08) | 50.0(0) | 75(0) | 1.19(0.10) | 0.85 |
| 18 | 78.0 | 91 | 0.77(0.05) | 78.9(1.2) | 90(1.1) | 0.77(0.07) | 0.22 |

the 16 distributions binned at every 0.05 seconds. Our method was able to capture a wide range of beat-length distributions shapes, from the flat distribution of subject 4 to the peakier distribution of subject 16. Fig. 4-2 also shows the distribution pair for subject 5's video, which the KS test labeled statistically dissimilar. Our distribution is more normally distributed than the ECG distribution. Subject 5's heart rate, 1.6 Hz, was the highest heart rate in our test set. At high heart rates the duration of each beat is small, meaning that variation between beats is also small. We believe that our method struggled because of this.

### 4.1.1 Motion Amplitude

Pulse motion constitutes only a part of total involuntary head movement. We quantify the magnitude of the different movements within $[0.75, 5]$ Hz by calculating root mean square (RMS) amplitudes of the feature point trajectories.

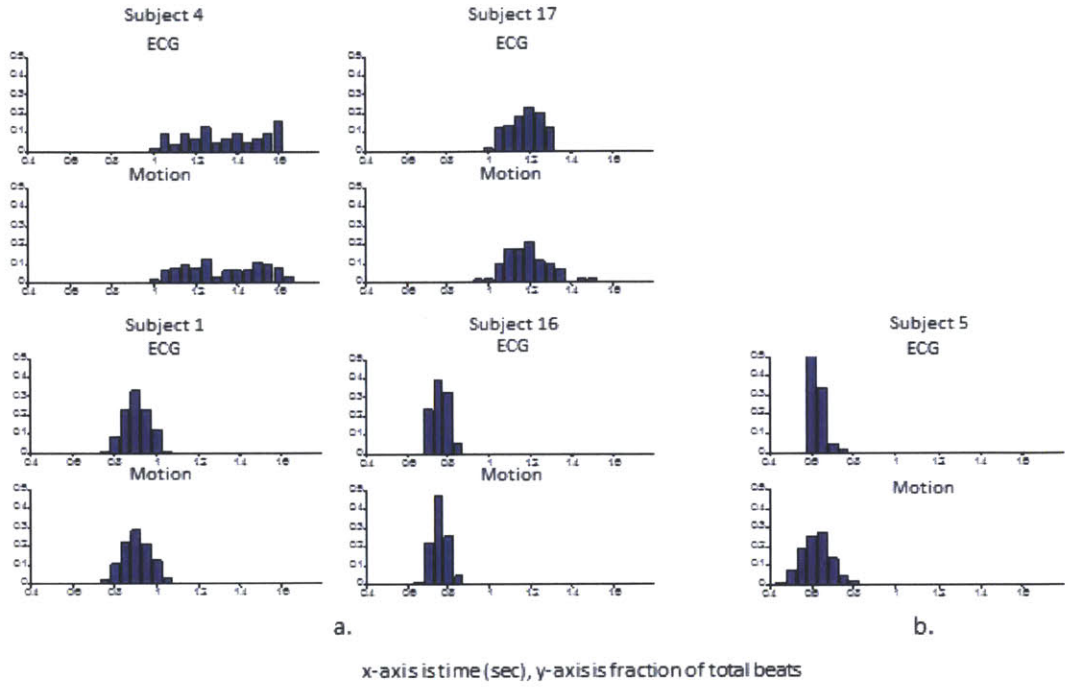



Figure 4-2: Beat distributions of the ECG and our motion method for subjects. (a) We were able to accurately capture a wide range of distribution shapes, such as for subjects 4, 17, 1 and 6. (b) We were not able to produce an accurate distribution for subject 5.

Table 4.2: Mean (std. dev.) RMS amplitudes in pixels for the $x$ and $y$ feature point trajectories for all subjects. Values are shown after filtering to within 0.05 Hz of the pulse frequency (RMS Pulse) and without filtering (RMS Total).

| | RMS Pulse | RMS Total | $\frac{\text{RMS Pulse}}{\text{RMS Total}}$ |
|---|---|---|---|
| $x$ | 0.17(0.11) | 0.44(0.05) | 0.36 |
| $y$ | 0.11(0.05) | 0.29(0.08) | 0.38 |

Table 4.2 presents the amplitudes averaged across all subjects, with the $x$ and $y$ trajectories averaged separately. We calculated the mean amplitude after filtering within 0.05 Hz of each subject's pulse frequency (RMS Pulse) and without filtering (RMS Total). The $x$ axis had larger pulse and total motion amplitudes, but the ratio between the two was slightly higher for the $y$ axis. For both the $x$ and $y$ axes the pulse amplitude was less than 40% of the total amplitude, indicating that factors other than pulse cause a majority of head motions when a person is sitting still - even when only considering the frequency band $[0.75, 5]$ Hz.

### 4.1.2 Noise Analysis


Figure 4-3: Frames from videos of subject 1 with varying levels of additive gaussian noise. The top row shows frames before filtering. The bottom row shows frames after bilateral filtering. Noise ranged from a standard deviation of 5 to 500 pixels. Bilateral filtering smooths images while retaining edges.

We performed noise analysis to evaluate the robustness of our system. We added zero-mean Gaussian noise to the videos, sweeping the standard deviation $\sigma_{noise}$ from 5 to 500 pixels. The noise was added to each channel independently. To smooth the noise we applied a bilateral filter [24] to each frame of the noisy videos. The bilateral filter is a smoothing filter that preserves edges and corners, which can be useful for tracking. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels, with weights depending on both Euclidean and intensity differences. Formally, the intensity of pixel $p$ is replaced by:

$$\frac{1}{W_p}\sum_q G_{\sigma_d}(||p-q||)G_{\sigma_r}(||I_p-I_q||)I_q \tag{4.1}$$

where $I_p$ is the current intensity at $p$, $I_q$ is the intensity at each pixel $q$, $G_{\sigma_d}$ and $G_{\sigma_r}$ are Gaussian functions depending on Euclidean and intensity differences, and $W_p$ is

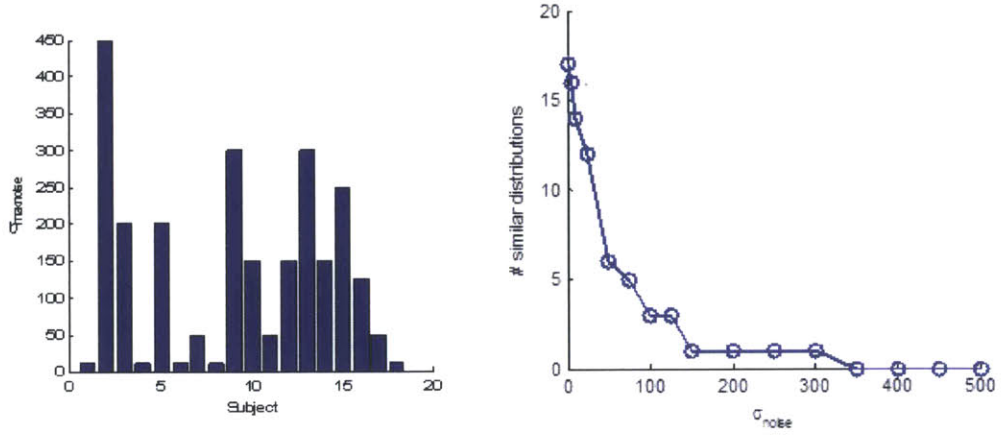
a normalization factor. We set $\sigma_d = 3$ and $\sigma_r = \sigma_{noise}$. Fig 4-3 shows frames from videos of subject 1 with different levels of noise before and after applying the bilateral filter.

For each subject we found $\sigma_{maxnoise}$, the maximum noise standard deviation before our method first produced an average pulse rate outside 5% of the true rate. Bilateral filtering resulted in a better (larger) $\sigma_{maxnoise}$ for 16 of the 18 subjects, worse performance for 1 subject, and no change in performance for 1 subject. Overall, filtering improves our method's performance against noise. Therefore, the rest of our results are compiled using bilateral filtering.

Fig. 4-4a presents $\sigma_{maxnoise}$ for each subject. There is a large variance in $\sigma_{maxnoise}$ across the subjects. Fig. 4-4b plots the number of distributions of beat lengths that are similar to ground truth (using the KS test) as $\sigma_{noise}$ is increased. Most of the 17 distributions that were initially found to be similar to ground truth are dissimilar when $\sigma_{noise} \approx 50$.

The large variance in $\sigma_{maxnoise}$ suggests that there are video-specific factors that affect performance. We hypothesized that the variance is caused by two video-specific factors: strength of head motion and feature point quality. A video with large and clear head motions related to cardiac activity should result in a higher $\sigma_{maxnoise}$. A video with feature points on good image patches, i.e. patches with high gradients, should result in less tracking error, which also leads to a higher $\sigma_{maxnoise}$. We tested these hypotheses by quantifying motion strength and feature point trackability with metrics. The metrics are based on the original videos (without added noise), and we test whether they have any correlation with $\sigma_{maxnoise}$.

We developed two metrics measuring motion strength. The first, which we call $\beta$, is the difference between the total energy of the feature points at $f_{pulse}$ to the maximum total power at any other frequency outside 0.05 Hz of the pulse frequency. $\beta$ captures the relative energy of pulse motion to the next strongest periodic motion. If the energy at the pulse frequency is barely larger or even smaller than the energy at another frequency, $\beta$ will be small. It is possible for $\beta$ to be negative, which means that the energy at another frequency is larger than the pulse frequency. As shown in




(a) (b)

Figure 4-4: Plots presenting results after adding varying levels of Gaussian noise to our videos. (a) shows $\sigma_{maxnoise}$, the standard deviation of the first noise level at which our method gave an incorrect average pulse rate. (b) shows the number of similar distributions reported by the KS test as noise is added to the videos.

Fig. 3-3 this is possible when the harmonic frequency is large or when there is strong periodic component corresponding to other body movement. The second metric, $\gamma$ is based on the PCA decomposition of the feature point trajectories of the original video. $\gamma$ is a weighted sum of the variance of each PCA signal $s_i$ with main frequency $f_i \approx f_{pulse}$:

$$\gamma = \sum_{i=1}^{5} \rho_i \lambda_i \mathbb{1}\{|f_{pulse} - f_i| < 0.05\,\text{Hz}\} \tag{4.2}$$

The weight for each $s_i$ is the periodicity metric $\rho_i$ that we defined earlier in Chapter 3. A signal with large amplitude (high $\lambda$) and large periodicity (high $\rho$) will result in a high $\gamma$ value. Fig. 4-5 shows our results comparing $\sigma_{maxnoise}$ to $\beta$ and $\gamma$. Both $\beta$ and $\gamma$ have a significant positive correlation with $\sigma_{maxnoise}$ (Pearson $R$ coefficients of 0.83, 0.76, $p$ values $< 0.01$). This suggests that the strength and clarity of the head motions are contributing factors to noise performance.

Next, we wanted to determine the effect of feature point quality on $\sigma_{maxnoise}$. We quantify feature point quality, $\kappa$, as the variance of the luminosity intensity within the 40 x 40 pixel window surrounding the point. This window size is the same size we

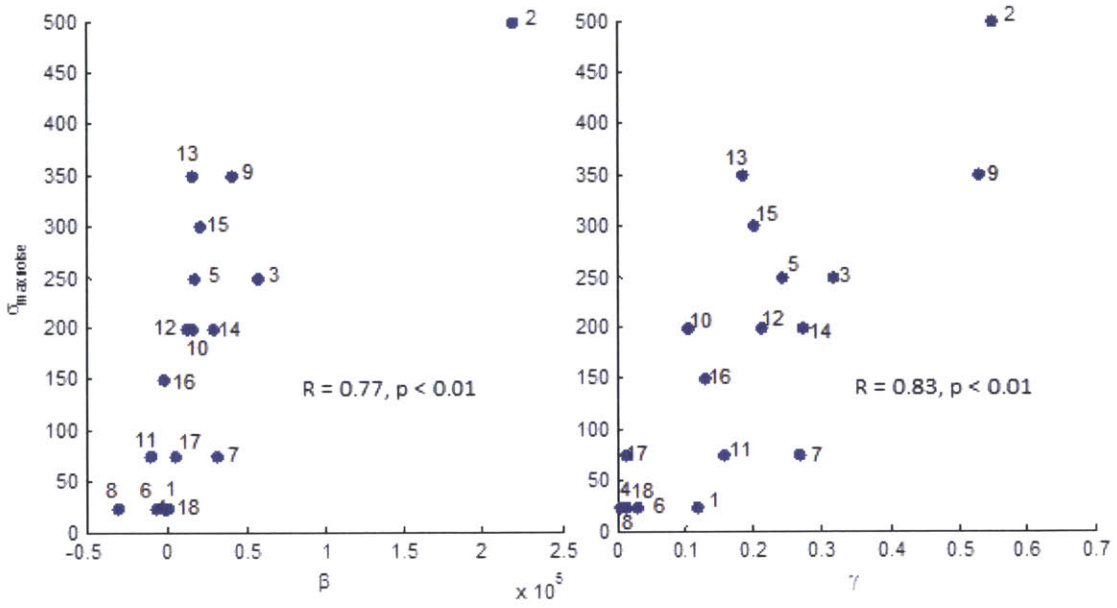
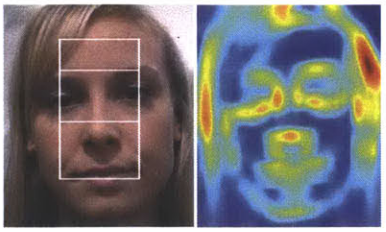



Figure 4-5: Plots of $\sigma_{maxnoise}$ vs. $\beta$ and $\gamma$ our two metrics measuring head motion strength. Both variables are significantly correlated with $\sigma_{maxnoise}$.

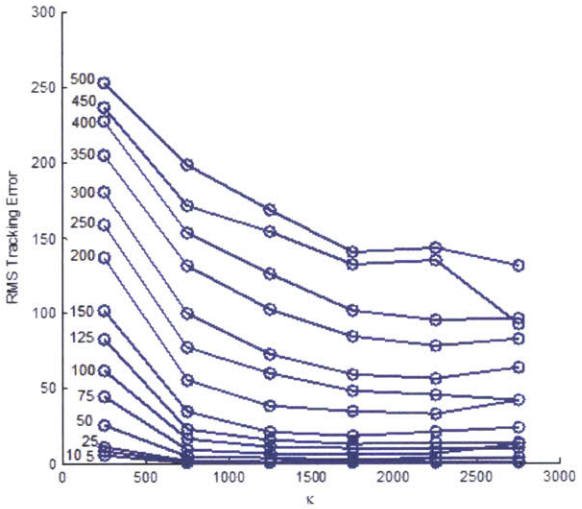
used for the Lucas-Kanade tracker. Intuitively, a low variance indicates a featureless or poor region. Fig. 4-6a shows a heatmap of $\kappa$ on a frame of subject 3's video, with red being the highest quality. Locations near the mouth and nose have high $\kappa$ while the cheeks and forehead do not.

First, we explored how $\kappa$ relates to tracking error. We measured the tracking error for each feature point on the videos. The tracking error at a frame is the distance between the point's location in the noiseless video and its location in the noisy video. Fig. 4-6b plots the average RMS tracking error of the feature point trajectories binned at different values of $\kappa$ for each noise level. We see a clear negative relationship between tracking error and $\kappa$ for all noise levels, as expected. However, at higher noise levels even the best points have quite a large tracking error, indicating that points are drifting in location over time. Despite this relationship between $\kappa$ and tracking error, we were unable to find any relationship between the average $\kappa$ for a video and $\sigma_{maxnoise}$ (see Fig. 4-6c).

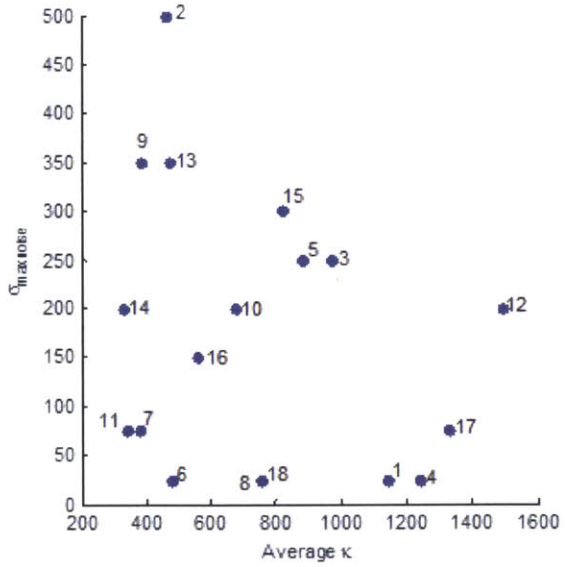
We ran a final experiment testing whether selecting a small subset of high quality feature points will result in a larger $\sigma_{maxnoise}$. We added a varying number of feature



(a) A heatmap of $\kappa$ for a frame of subject 3's video. Hotter colors (e.g. red) indicate high $\kappa$ while coolor colors (e.g. blue) indicate low $\kappa$. The best $\kappa$ within the feature point region is near the nose and mouth.




(b) Average root mean square tracking error of feature points for all subjects binned at different $\kappa$ values. Each line corresponds to a different noise level ($\sigma_{noise}$ is indicated next to each line). Larger $\kappa$ results in lower tracking error for all noise levels.




(c) Plot of $\sigma_{maxnoise}$ vs. average $\kappa$ in a video. The variables are not significantly correlated.

Figure 4-6: Analysis of $\kappa$, our metric measuring feature point quality.

points with high $\kappa$ and recorded $\sigma_{maxnoise}$. We swept $N$, the number of points, from 150 to 1500 in increments of 150. Fig. 4-7 plots the average $\sigma_{maxnoise}$ value across all subjects for each level. For comparison, we also show results when only selecting points with low $\kappa$ and selecting points randomly within the region of interest. For the random selection we selected 10 random sets of feature points and record the average $\sigma_{maxnoise}$. The results of this experiment show that selecting points randomly and by high $\kappa$ yields similar results. Selecting points with low $\kappa$ performed predictably poorly. Adding more feature points increased performance in all scenarios but helped low $\kappa$ the most. Increasing the number of feature points from 150 to 1500 only increased $\sigma_{maxnoise}$ by 30 when selecting points randomly or by high $\kappa$.

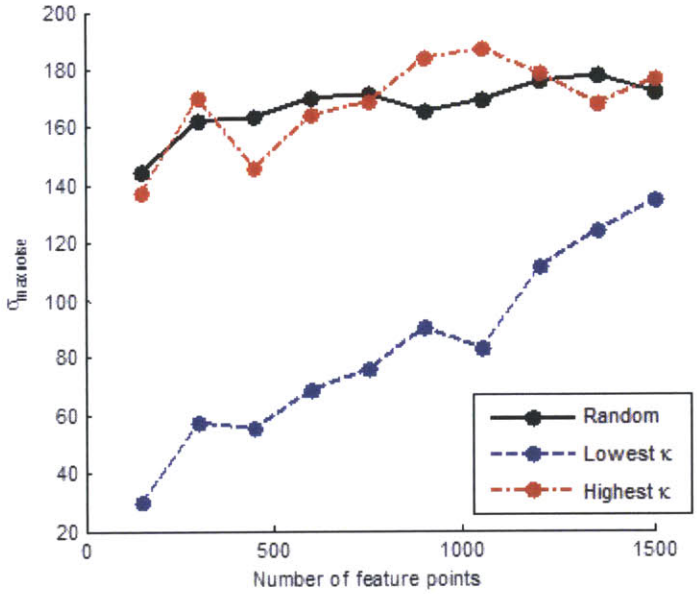



Figure 4-7: The average $\sigma_{maxnoise}$ of all subjects vs. the number of feature points selected. 1500 points is the number of feature points used in the original results (see Fig. 4-4a). We selected points in three ways: randomly around the region of interest, by the highest $\kappa$ and by the lowest $\kappa$. The random results were repeated 10 times and averaged. Random selection and the high $\kappa$ selection peformed similarly while low $\kappa$ yielded the worst results. Adding more feature points increased $\sigma_{maxnoise}$ for all sampling methods but helped the low $\kappa$ method the most.

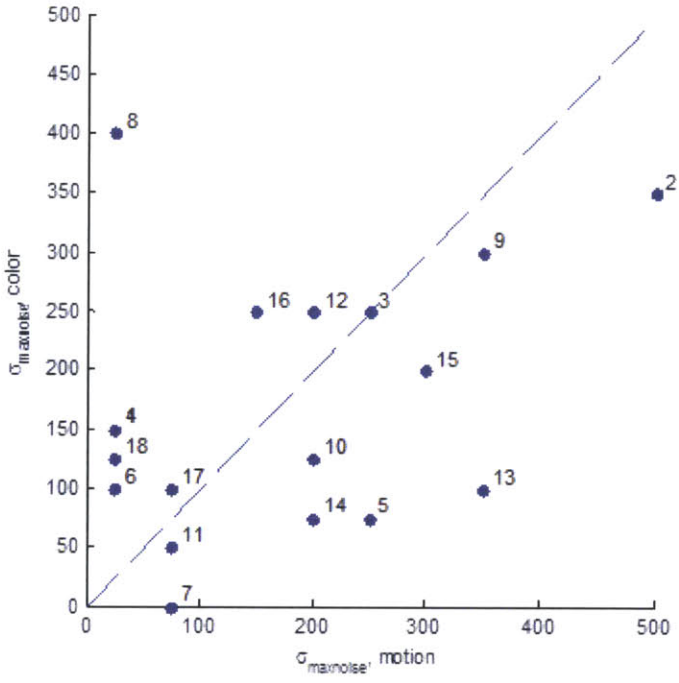



Figure 4-8: Comparison of our method to color-based detection. $\sigma_{maxnoise}$ is the maximum noise standard deviations where either color and motion give reliable results. Our method worked longer (under the blue dotted line) for 9 of the 18 subjects while color worked longer for 8 subjects. The color method failed to give a correct result for subject 7 before the addition of noise.

### 4.1.3 Comparison to Color-Based Detection

We compare the robustness of our method to a color-based pulse detection system [18] in the presence of noise. This color-based method spatially averages the R, G, and B channels in the facial area and uses independent component analysis (ICA) to decompose the signals into 3 independent source signals. The source with the largest peak in the frequency spectrum is then chosen as the pulse signal. We calculated $\sigma_{maxnoise}$ in an identical manner for the color method. Fig. 4-8 plots the results. Our method outperformed color for 9 of the 18 subjects, and performed worse for 8 subjects. Note that color failed to produce a correct pulse rate for subject 7 before adding any noise. A possible reason for this is that this video was overexposed. We found that the $\beta$ and $\gamma$ values are positively related to $\sigma_{maxnoise}$ for the color method,

similar to the earlier findings for motion. To our surprise we saw no relationship between color performance and skin tone.

## 4.2 Other Views of the Head

Figure 4-9: Reference frames from two videos of the back of the head and one of a face covered with a mask.

One of the advantages of motion-based detection over color is that a direct view of the skin is not needed. We took videos of the backs of the heads of 11 subjects and a video of one subject wearing a mask, as shown in Fig. 4-9. We were able to get average heart rates close to the true values for all videos, although the signal-to-noise ratio was less than for the frontal videos. One reason for this is that the back of the head is less trackable than the front due to a lack of distinct features.

## 4.3 Sleeping Newborns

We also tested our system on 30-60 second videos of three newborns recorded in situ at a hospital nursery (see Fig. 4-10). The videos were shot under varying lighting conditions while the babies were sleeping. We hand-picked these video sequences to have minimal body movements like turning over, yawning, stretching, etc. We obtained videos of the babies' actual pulse rates from hospital-grade monitors measuring the perfusion of blood to the skin. Our algorithm extracts pulse rates identical to those reported by the monitors.

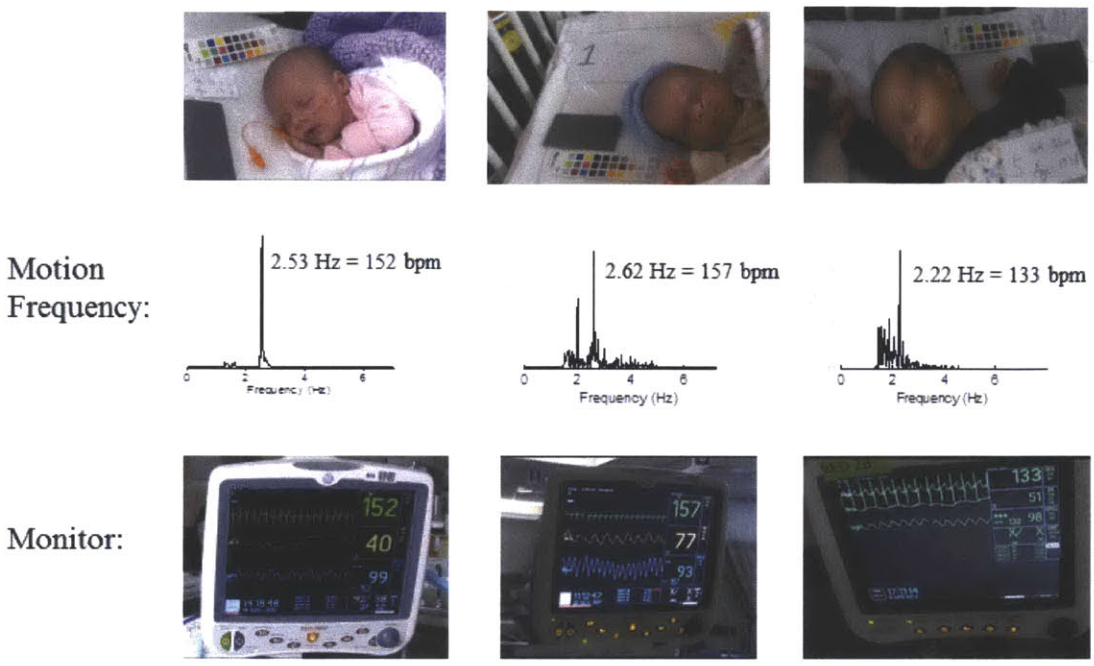



Figure 4-10: Results from videos of sleeping newborns. Our method produces pulse rates matching the actual bedside monitors.

## 4.4 Discussion

Our results show that it is possible to consistently obtain accurate pulse rate measurements from head motion. The results for beat detection were equally encouraging. Most of our beat interval distributions looked qualitatively similar to the ECG distributions, indicating that we do capture physiological variability. For 17 of the 18 subjects, we found that there was not a statistically significant difference between the ECG and the motion beat intervals. It is worth noting that this is a stronger test than is required in most clinical contexts. Typically heart rate variability (HRV) is used to dichotomize patients into high and low risk groups, so the precise shape of the distribution is not relevant. The relevant test would be whether the distribution of motion-generated intervals yields the same set of high risk individuals as ECG generated intervals. Since all of our subjects were healthy volunteers, we were not able to perform this test.

Several factors affected our results. First, our camera has a sampling rate of

30Hz. ECG used for HRV analysis normally has a sampling rate of at least 128 Hz. Cubic interpolation of our signal only partially addresses this discrepancy. Second, the variable and suboptimal lighting conditions can affect our feature tracking. We believe this to be the case for several of our videos. Finally, our videos were a maximum of 90 seconds long. Normally, HRV measures are computed over many hours to obtain reliable estimates.

# Chapter 5

# Other Applications of Our Method

While the main focus of our work is on detecting pulse from head motions, we were also able to apply our methods to other tasks.

## 5.1 Pulse From Other Body Parts

Blood circulation causes more than just the head to move in the human body. Arteries, veins and limbs are a few of the other body parts that move in synchrony with the pulse rate. By measuring and relating the movements of these different locations together we can understand more about a person's cardiovascular condition.

We tried getting a pulse rate from the carotid artery and the chest of one subject. Carotid pulsation can be felt with a finger below the jaw bone and it is quite common for doctors to measure pulse rate by feeling the pulsations of this artery. The chest moves in a different manner, vibrating to the movements of the heart. By placing a hand over the chest, one can feel the S1 and S2 sounds, better known as "lub-dub", during a full cycle.

Fig. 5-1 shows signals we extracted from the carotid artery and the chest. These signals were produced using the method we used for measuring head motions. Both signals have two peaks per cycle, which is a phenomenon we did not notice with most of the head motion signals. We hypothesize that this is related to the dicrotic notch, a secondary upstroke corresponding to the transient increase in aortic pressure upon

closure of the aortic valve.

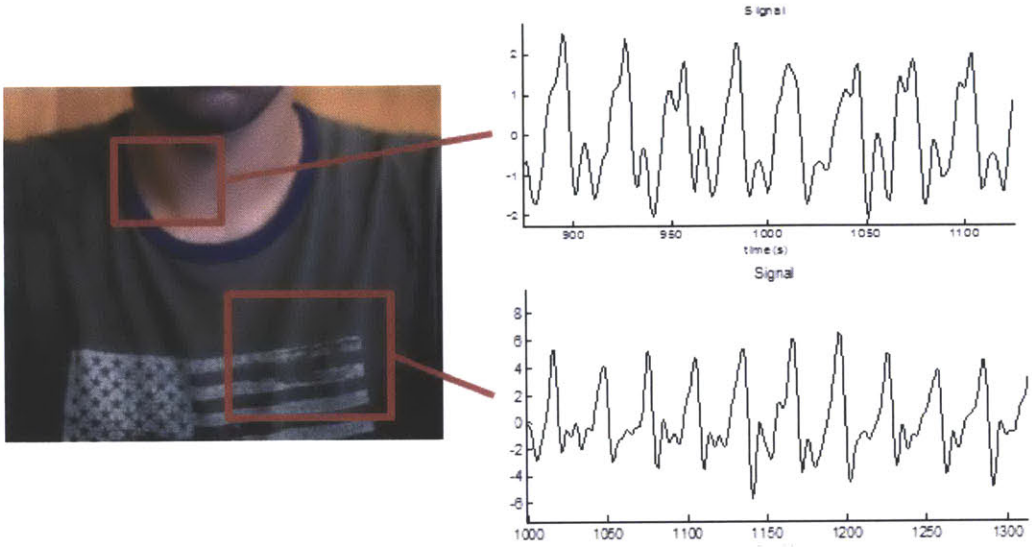



Figure 5-1: Pulse signals from the chest and carotid artery.

## 5.2 Respiration

Our methods are not limited to cardiac pulse. We are also able to measure respiration by tracking movements in the face and chest areas. All that is required to measure respiration using our method is to change the temporal filter passband applied to the feature point trajectories. Adult resting respiration signals are in a frequency range of $0.10 - 0.50$ Hz (9 - 30 cycles per minute) while infants have a higher range, between $0.50 - 1.50$ Hz. Fig. 5-2 shows examples of three signals from 2 of our adult subjects and one infant. The rates for the two adult subjects were quite low (around 8 - 10 seconds per cycle), likely because they were told to sit still as possible for the videos. The respiration rate for the infant matched real-time rates reported by a hospital monitor. Hospital respiration monitoring often requires a nasal cannula and is often unreliable. Our method provides an alternative that could be less irritating.

Finally, the respiration signals we extracted are cleaner than the signals we obtained for pulse, probably because respiration has a much larger amplitude of motion.

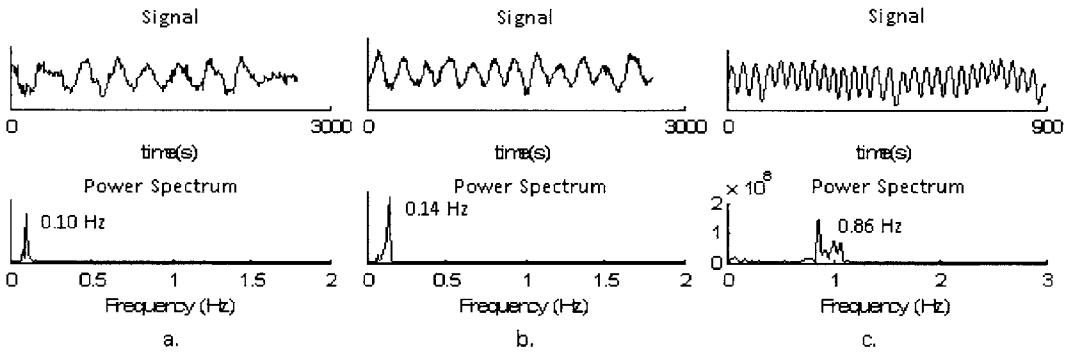



Figure 5-2: Respiration signals from 2 subjects (a,b) and a neonatal infant (c). The rates for the two adult subjects were quite low, likely because they were told to sit still as possible for the videos. The rate for the infant is near a normal adult's pulse rate, which is typical for newborns.

# Chapter 6

# Visualization Techniques

So far we have presented and evaluated a method of automatically extracting cardiovascular information from head movement. However, there is also a need for better visual aids that can provide insights in a more intuitive way. Visualization techniques in this area of research are important for several reasons. First, they can serve as tools to help understand the physiology behind the signal of interest. For example, by magnifying head motions we were able to understand the dominant axes of motion for different subjects. Second, visualization often helps open new research paths. Finally, visualization could be of clinical utility on its own by serving as an aid to understanding physiology without needing to collect precise measurements. In this chapter we present several visualization methods that we have developed. We describe these techniques in the context of head motions, but we believe that they can be used in other applications.

## 6.1 Single Images

The following two methods display motion information on a single or few images.

### 6.1.1 Eigenvector Plot

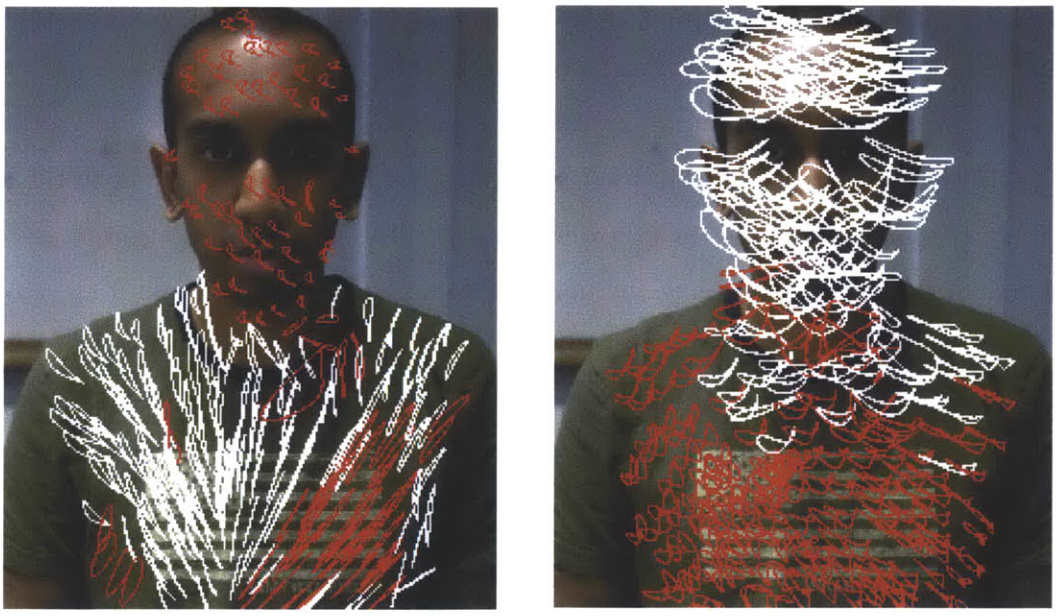
We have already shown an example of eigenvector plots in Fig. 3-4. We visualize each PCA signal of the head motion by creating an image of the eigenvector overlayed on a frame of the video. For each point we plot the $x$ and $y$ weights returned by the PCA algorithm as an arrow located over the point. Since most of the head motion is captured by the first few components, only a few images are needed to describe the directions of motion within a frequency range. These are also critical to understanding what the motion corresponding to our method's extracted pulse signal looks like.

### 6.1.2 Frequency Trajectory Plots

We developed a second image visualization technique that shows how points move at specific frequencies. The DFT of a signal returns a sinusoidal decomposition with a unique sinusoid for each frequency component. We reconstruct the motion at specific frequencies of our trajectories by obtaining their DFT coefficients and summing the sinusoids at those frequencies. To produce visualizations we plot the trajectories on the image of the video. Code A.1 in Appendix A shows MATLAB code for drawing a plot given the trajectories $\mathbf{X}, \mathbf{Y}$, the video frame to draw on and a vector of frequencies.

Fig. 6-1 shows examples of plots for one video of a subject at different frequencies. The color of a trajectory specifies the direction the point moves on the trajectory (clockwise for red, counterclockwise for white). Fig. 6-1a shows trajectories, magnified by 100, when considering respiration frequency and its first harmonic frequency. The ellipses from the left and right chest point away from each other since the chest expands and contracts when breathing. Assymetries of the trajectories between the left and right chest may be attributed to the assymetric lighting in the scene (the left side of the body is shaded). Fig. 6-1b is a plot at the pulse frequency and its first harmonic, magnified 1000 times. An interesting observation is that the head moves very differently from the chest at this frequency. First, the head ellipses are larger meaning that the head moves more. The head also seems to move as one uniform

unit separate from the chest. The head is moving mostly in the horizontal axis while the motions on the chest are both horizontal and vertical. These observations may suggest that the force acting on the head is not the same as the force on the chest. Or, the head moves differently because of its mechanical properties. In any case these plots could provide valuable insights about the precise causes of pulsatile head motion, a question that is still unanswered (see discussion on this topic in Chapter 2).



(a) (b)

Figure 6-1: Frequency trajectory plots. Red trajectories are points moving clockwise and white are counterclockwise. (a) is at the respiration frequency magnified 100 times. The ellipses from the left and right chest point away from each other since the chest expands and contracts when breathing. (b) is a plot at the pulse frequency magnified 1000 times. The head moves more than the chest in this case.

## 6.2 Video Motion Amplification

Our second set of tools create rerendered videos with motions amplified, allowing a user to more easily see the movements.

## 6.3 Basic Amplification

We track feature points on the head as described in Chapter 3, filter their trajectories to a band of interest and amplify changes in position from frame to frame. We make the magnified video by applying a Delaunay Triangulation to the points and warping each triangle from the input video to fit the new, amplified trajectories.

Assume we obtain the horizontal and vertical trajectories of each feature point $i = 1 \cdots N$ as matrices $\mathbf{X}$ and $\mathbf{Y}$, where $\mathbf{X}_{i,j} = x_i(j)$ and $\mathbf{Y}_{i,j} = y_i(j)$. First, we filter these trajectories to only contain frequencies within $[0.75, 5]$ Hz. Next, we create amplified trajectories **Xa** and **Ya** (see Code A.2 in Appendix A).

Using **Xa** and **Ya** we can re-render the input video with head motions in the band $[0.75, 5]$ Hz magnified. To do so we first apply a Delaunay triangulation to the first frame of the input video using the point locations in $\mathbf{X}(1 \rightarrow N, 1)$ and $\mathbf{Y}(1 \rightarrow N, 1)$. Fig. 6-2 shows an example of applying a triangulation on the first frame of subject 5's video (with many feature points removed for clarity). The remaining frames of the video are rerendered by affine warping each triangle using the trajectories **Xa** and **Ya**. Fig. 6-3 shows an example of a magnified sequence of frames. Notice how both the up/down and left/right axes of motion are clearly visible after applying our magnification. The movements are not visible in the normal frame sequence.

## 6.4 Typical Beats

A recording of a subject will almost always span several cardiac cycles. It can be useful to have a simple model of the head motions observed over all the cycles. We developed a way to extract the "typical motion" of a subject's head per cardiac cycle in a given video. One could more easily understand an individual's head motion and inter-subject differences by comparing the typical motions.

We first run our signal extraction method to obtain the trajectories of feature points and the beat intervals from the pulse signal. We use dynamic time warping (DTW) to average the trajectories from the intervals together. The resulting averaged

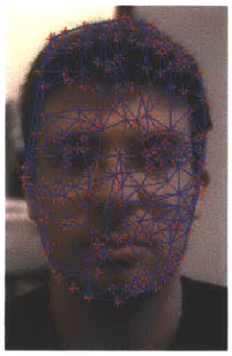


Figure 6-2: Example of applying a Delaunay triangulation to feature points on subject 5's face. Red stars denote feature points and the blue lines show the triangulation between the points.

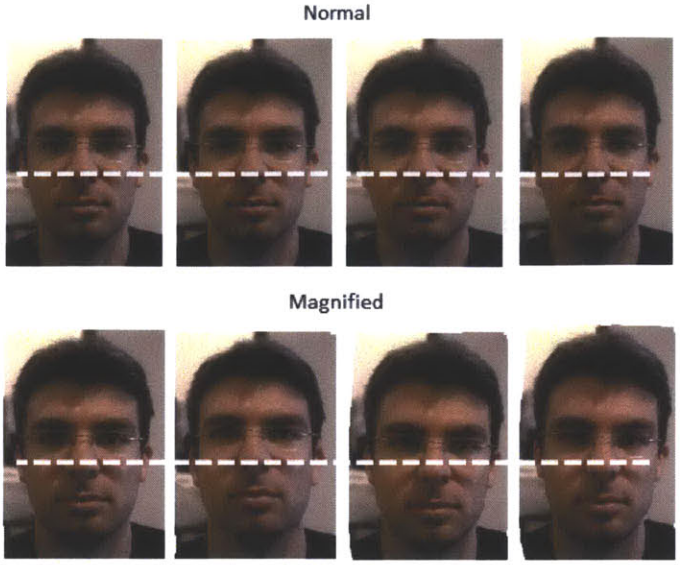



Figure 6-3: A comparison of a sequence of frames before and after applying motion magnification. The head motion is much more visible after magnification.

trajectory is the typical head motion trajectory for this subject's video. It is important to remember that the typical motion we extract is for a given video and may not generalize to the same subject in a different environment or posture.

### 6.4.1 Input

As described in Section 6.3, we store filtered feature point trajectories in matrices $\mathbf{X}$ and $\mathbf{Y}$. We also store beat intervals in $\mathbf{K}$, in which each row defines one interval. For row $i$ the element $\mathbf{K}_{i,1}$ is the sample number at which the interval starts and $\mathbf{K}_{i,2}$ is the sample at which the interval ends. We assume that the samples in $\mathbf{K}$ are based on the same sampling rate used for $\mathbf{X}$ and $\mathbf{Y}$, i.e. 30 Hz.

### 6.4.2 Algorithm

Given $\mathbf{X}, \mathbf{Y}$ and $\mathbf{K}$, we segment the video into intervals and compute the average trajectory per interval. First, we discard intervals from $\mathbf{K}$ that are greater than 25% from the mean interval length.

Because each interval will be of a slightly different length it is first necessary to time-align the signals from all the intervals before taking an average. We use dynamic time warping or DTW to do this. DTW is a method of aligning two signals and/or measuring similarity between two signals which may vary in time or speed. The algorithm is called "dynamic" because it makes use of dynamic programming to store the current best match between the signals.

We first select the median beat interval in length. The trajectories from the rest of the intervals will be aligned to the median interval's trajectory. DTW does this by assigning a distance between each alignment. The distance in our case is simply the Euclidean distance between the trajectories. Code A.3 in Appendix A presents MATLAB code for extracting the typical trajectory **Xtyp** and **Ytyp** given $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{K}$ and $m$, the row index of $\mathbf{K}$ corresponding to the median beat interval.

Using **Xtyp** and **Ytyp** we can create a magnified video of the motion using the algorithm presented in the last section, or we can analyze the signals directly. Fig. 6-4 shows a simplified representation of the typical motions of our 18 subjects. The curves in these plots are the typical $x$ and $y$ trajectories per feature point (instead of the high-dimensional position of all points together). The trajectories for different subjects are not aligned in time since our pulse signals are not aligned to a specific physiological event. Subjects have different trajectory lengths due to different pulse rates. Ignoring the differences in trajectory lengths, the morphology of the motions still varies widely across our subject pool. However there are subsets of subjects with similar patterns. For example, the curves for subjects 17 and 18 have a similar morphology as do subjects 9 and 16. A future direction with this work would be to link our motion signals to a physiological phenomena so that we can quantify the differences between the typical motions. Measuring differences is the first step

in differentiating between abnormal and normal motion patterns, which may have clinicial significance.

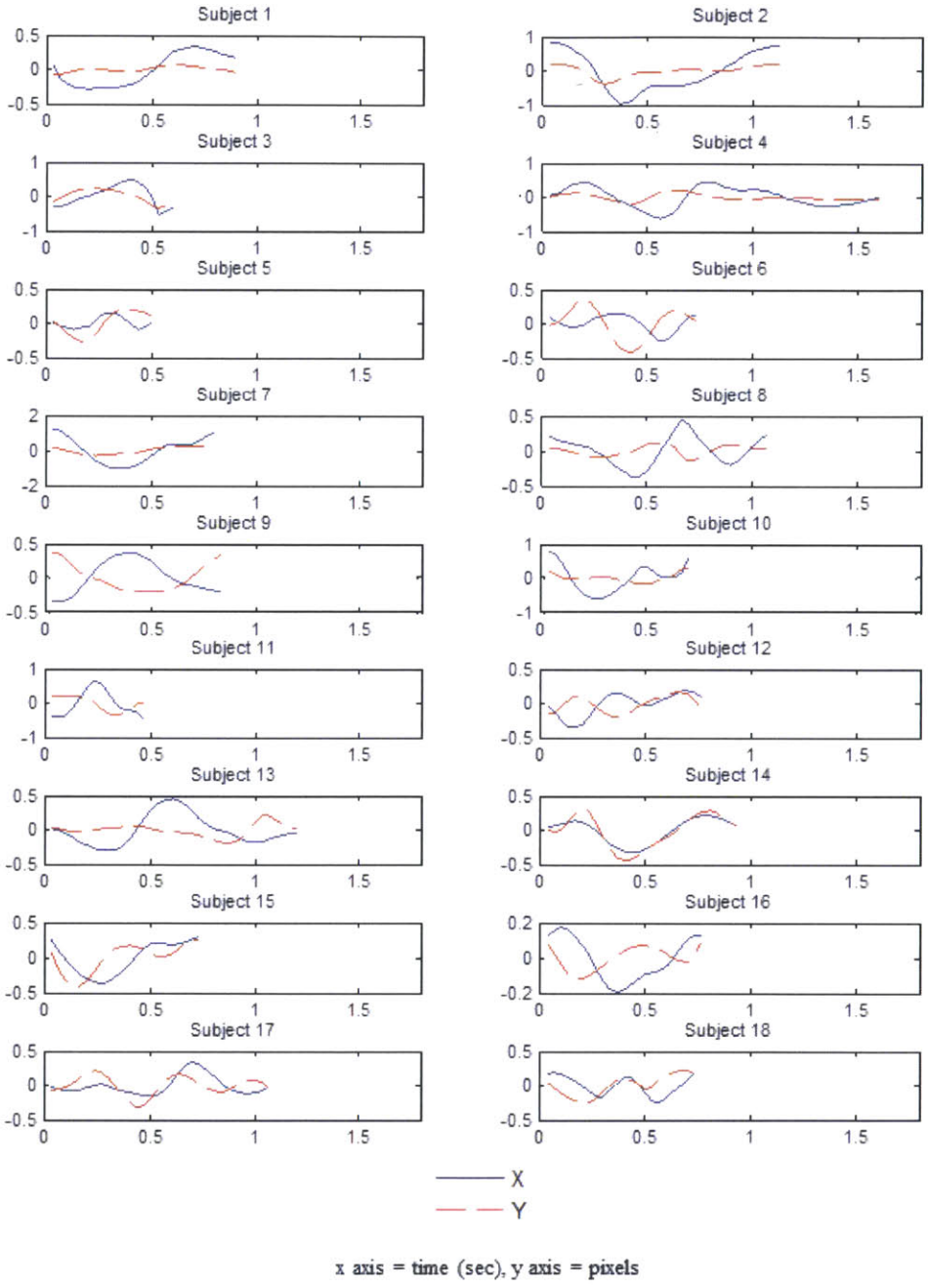



Figure 6-4: The typical $x$ and $y$ feature point trajectory for each subject.

# Chapter 7

# Summary & Conclusion

In this thesis we described a novel approach that offers a non-contact means of cardiac monitoring. Our method takes video as input and extracts heart rate and beat measurements from the head motion associated with the pumping of blood at each heartbeat. A combination of feature tracking, frequency filtering and PCA allows us to identify the component of motion corresponding to the pulse, and we then extract peaks of the trajectory to identify individual beats. When evaluated on regular video recordings of 18 still subjects with a view of the face, our method produced virtually identical heart rates to an ECG and produced similar beat-length distributions for 17 subjects. Our method also produced correct rates for videos of the back of the head, when the head is occluded by a mask, and for infants sleeping in a hospital. From these results we conclude that pulsatile head motion is large enough to be captured from a consumer camera, and can be separated from other head motions to produce a reliable pulse signal that can yield a pulse rate and variability measures.

We also ran experiments to understand how our method degrades with the addition of noise. We applied zero-mean Gaussian noise to the 18 face videos and found that the noise level at which our method produced an incorrect average pulse rate was positively correlated to the strength and clarity of the head motion in the given video. Performance was not correlated with the texture of the face, which surprised us. We conclude that strength of the head motion is the dominant factor determining how well our method performs with noise. However we do not yet have a complete

understanding of why some videos work better than others under noisy conditions. When we develop a better theory for this, we can then make our method work better in a variety of real-world conditions.

We also presented methods for visualizing the head motions including motion magnification, typical beat magnification and ellipse plots to show movement in a single image. These visualization tools can help a clinician make subjective decisions on a person's health without needing precise measurements. They also serve as exploratory tools to guide researchers developing quantitative methods. A future direction is to develop a set of standard visualization tools for different medical applications and make them work in real-time.

Finally, we showed that the methods we propose are useful for other applications besides extracting pulse from head motion. Initial results demonstrate that we can get pulse from other body parts such as the chest and carotid artery. This finding opens future research into combining signals around the body to obtain deeper physiological information. For example, by measuring the timing difference between movements in the chest, carotid and hands we might get a good estimate for blood velocity. We also demonstrated that we can measure respiration from head/chest movement by merely changing the frequency band of interest. Respiration is usually monitored in hospitals using a nasal cannula, and is often unreliable. Our method provides an alternative that may be more practical.

In this work we considered the frequency and variability of the pulse signal. However, head movement can offer other information about the cardiac cycle. If head displacement is proportional to the force of blood being pumped by the heart, it may serve as a useful metric to estimate blood stroke volume and cardiac output (the product of blood volume and pulse rate). Additionally, the direction of the movement might reveal asymmetries in blood flow into or out of the head. This might be useful for the diagnosis of a stenosis, a partial blockage, of the carotid arteries. We would like to perform a study with patients in a hospital to see whether our methods have diagnostic value for cardiac conditions.

Our experiments were done with still subjects. Another future direction is to make

our method work when subjects are moving. Considering that our method already overcomes extraneous involuntary head motions to extract pulse, we believe that it is possible to subtract out higher amplitude motions such as talking or changes in posture. However, this may require more sophisticated feature tracking and decomposition methods.

Finally, as always in the healthcare space, it is important to be able to interpret the signals we are working with. Our method produces a new physiological signal that may have clinical uses. However we still do not have a good understanding of where the force causing the head to move is coming from. In Chapter 2 we derived simple estimates of the acceleration of the head due to aortic and carotid blood flow forces. These calculations indicate that the aortic force causes a majority of the acceleration of the head. However we are not certain of the accuracy of our simple physical model and more work needs to be done to verify this.

# Appendix A

# Visualization Code

Code A.1: Drawing ellipses for visualization.

```
function drawEllipses(X,Y,frame,frequencies)
%X,Y: N x T matrices of trajectories
%frame: frame of video to plot the trajectories
%frequencies: a vector of frequencies to consider.

N = size(X,1);
T = size(X,2);

%Get the frequency indices in the DFT.
fr = 30; %sample rate
f = fr * linspace(0,1,T);
fidx = [];
for i=frequencies
    [¬,idx] = min(abs(f - i);
    fidx = [fidx idx];
end

figure; hold on;
imshow(frame);

t = 1:T;
```

```
for i = 1:n
    XFC = fft(X(i,:));
    YFC = fft(Y(i,:));

    for j=fidx
        xS = xS + abs(XFC(i,j))/N * cos(2*pi*(j-1)/N * t + ...
            angle(XFC(i,j)));
        yS = yS + abs(YFC(i,j))/N * cos(2*pi*(j-1)/N * t + ...
            angle(YFC(i,j)));
    end

    %Compute direction of movement using right-hand rule.
    dir = (xS(2)-xS(1))*(yS(3)-yS(2))-(yS(2)-yS(1))*(xS(3)-xS(2)) ...
        > 0;
    if(dir == 1)
        plot(xS,yS,'r');
    else
        plot(xS,yS,'w');
    end
end
end
```

Code A.2: Amplifying Feature Point Trajectories

```
function [Xa,Ya] = amplifyTrajectories(X,Y,m)
%X,Y: N x T matrices of trajectories
%m is magnification factor.
Xa = zeros(N,T);
Ya = zeros(N,T);
Xa(:,1) = X(:,1);
Ya(:,1) = Y(:,1);
for t = 2:T
    Xa(:,t) = X(:,t-1) + m * (X(:,t) - X(:,t-1));
        Ya(:,t) = Y(:,t-1) + m * (Y(:,t) - Y(:,t-1));
end
```

Code A.3: Producing Typical Beat Trajectories

```
function [Xtyp,Ytyp] = typicalMotion(X,Y,K,m)
%X,Y: N x T matrices of trajectories
%K: matrix defining the pulse intervals.
%m: the index in K of the interval to which we want to warp the ...
    rest of the intervals to.
N = size(X,1);
L = size(K,1);

A = [X;Y]; %Stack X and Y into matrix A for ease
Am = A(:,K(m,1):K(m,2));
Atyp = zeros(2*N,1:K(m,2)-K(m,1)+1);

for i = 1:L
    Ai = A(:,K(i,1):K(i,2));
    %DTW returns Ti warped in time to match Tm as closely as ...
        possible.
    Atyp = Atyp + DTW(Ai,Am);
end

Atyp = 1/L * Atyp;
%Return the X and Y typical trajectories from Atyp
Xtyp = Atyp(1:N/2,:);
Ytyp = Atyp(N/2+1:end,:);
end
```

# Bibliography


[1] L. Adde et al. Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study. *Developmental Medicine & Child Neurology*, 52:773–778, 2010.

[2] G. Bonmassar et al. Motion and ballistocardiogram artifact removal for interleaved recording of eeg and eps during mri. *NeuroImage*, 16:1127–1141, 2001.

[3] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[4] SW. Chen et al. Quantification and recognition of parkinsonian gait from monocular video imaging using kernelbased principal component analysis. *BioMedical Engineering OnLine*, 10, 2011.

[5] M. Delano. A long term wearable electrocardiogram (ecg) measurement system. *Master's Thesis, Massachusetts Institute of Technology*, 2012.

[6] M. Garbey et al. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Trnas Biomed Eng*, 54(8):1418–1426, 2007.

[7] E. Greneker. Radar sensing of heartbeat and respiration at a distance with applications of the technology. *Proc. Conf. RADAR*, pages 150–154, 1997.

[8] D. He. A continuous, wearable, and wireless heart monitor using head ballistocardiogram (bcg) and head electrocardiogram (ecg). *Conf Proc IEEE Eng Med Biol Soc. 2011*, 2011, 2011.

[9] D. Holdsworth et al. Characterization of common carotid artery blood-flow waveforms in normal human subjects. *Physiol. Meas.*, 20:219–240, 1999.

[10] O.T. Inan et al. Robust ballistocardiogram acquisition for home monitoring. *Physiological Measurement*, 30:169–185, 2009.

[11] K.K. Kim et al. A new method for unconstrained pulse arrival time measurement on a chair. *J. Biomed. Eng. Res.*, 27:83–88, 2006.

[12] C. Liu et al. Motion magnification. *ACM Trans. Graph.*, 24:519.

[13] G. Lu et al. Contact-free measurement of heartrate variability via a microwave sensor. *Sensors*, 9:9572–9581, 2009.

[14] P. Lynch. Human head anatomy with external and internal carotid arteries. http://www.flickr.com/photos/patrlynch/450142019/, April 2007.

[15] U. Morbiducci et al. optical vibrocardiography: A novel tool for the optical monitoring of cardiac activity. *Ann. Biomed. Eng.*, 35:45–48, 2007.

[16] M. Pediaditis et al. Vision-based human motion analysis in epilepsy - methods and challenges. *Proceedings of IEEE ITAB*, pages 1–5, 2010.

[17] Philips. Philips vitals signs camera. http://www.vitalsignscamera.com, 2011.

[18] M.Z. Poh et al. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774, 2010.

[19] HN. Sabbah et al. Noninvasive evaluation of left ventricular performance based on peak aortic blood acceleration measured with a continuous-wave doppler velocity meter. *Circulation*, 74, 1986.

[20] I. Starr et al. Studies on the estimation of cardiac output in man, and of abnormalities in cardiac function, from the hearts recoil and the bloods impacts; the ballistocardiogram. *The American Journal of Physiology*, 127:1–28, 1939.

[21] S. Suzuki et al. Development of non-contact monitoring system of heart rate variability (hrv)-an approach of remote sensing for ubiquitous technology. *Proc. Int. Conf. Ergonom. Health Aspects Work Comput.*, pages 195–203, 2009.

[22] Z. Syed et al. Computationally generated cardiac biomarkers for risk stratification after acute coronary syndrome. *Sci Transl Med*, 3(102):102ra95, 2011.

[23] K.S. Tan et al. Real-time vision based respiration monitoring system. *Proceedings of CSNDSP*, pages 770–774, 2010.

[24] C. Tomasi. Bilateral filtering for gray and color images. *Proceedings of ICCV*, pages 839–846, 1998.

[25] S. Ulyanov and V. Tuchin. Pulse-wave monitoring by means of focused laser beams scattered by skin surface and membranes. *Proc. SPIE*, pages 160–167, 1993.

[26] W. Verkruysse et al. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008.

[27] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of IEEE Conference on Computer Vision and Patern Recognition*, pages 511–518, 2001.

[28] CW Wang et al. Vision analysis in detecting abnormal breathing activity in application to diagnosis of obstructive sleep apnoea. *Proceedings of IEEE Eng Med Biol Soc*, pages 4469–4473, 2006.

[29] F.P. Wieringa et al. Contactless multiple wavelength photoplethysmographic imaging: a first step toward spo2 camera technology. *Ann. Biomed. Eng.*, 33(8):1034–1041, 2005.

[30] H. Wu et al. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proceedings SIGGRAPH 2012)*, 31(4), 2012.