Vital Sign Estimation from Passive Thermal Video

Ming Yang[†] , Qiong Liu[‡], Thea Turner[‡], Ying Wu[†]
[†]Dept. of EECS, Northwestern Univ.
2145 Sheridan Rd., Evanston, IL 60208
{mya671,yingwu}@ece.northwestern.edu
[‡]FX Palo Alto Laboratory, Inc.
3400 Hillview Ave., Palo Alto, CA 94304
{liu,turner}@fxpal.com

Abstract

Conventional wired detection of vital signs limits the use of these important physiological parameters by many applications, such as airport health screening, elder care, and workplace preventive care. In this paper, we explore contact-free heart rate and respiratory rate detection through measuring infrared light modulation emitted near superficial blood vessels or a nasal area respectively. To deal with complications caused by subjects' movements, facial expressions, and partial occlusions of the skin, we propose a novel algorithm based on contour segmentation and tracking, clustering of informative pixels, and dominant frequency component estimation. The proposed method achieves robust subject regions-of-interest alignment and motion compensation in infrared video with low SNR. It relaxes some strong assumptions used in previous work and substantially improves on previously reported performance. Preliminary experiments on heart rate estimation for 20 subjects and respiratory rate estimation for 8 subjects exhibit promising results. [1]

1. Introduction and background

Human heart and respiratory rates are important vital signs for heath monitoring. Conventionally, the heart and respiratory rates are measured by attaching sensors to a human body that are wired to preamp and processing instruments, *e.g.* Piezo Pulse Transducer and Electro-CardioGraphy (ECG) electrodes for heart rate, and Piezo Respiratory Transducer for respiratory rate measurement. These wired measurement sensors place severe restrictions on applications using these two parameters. Doppler ultrasound, photoplethysmopgraphy (PPG), and laser doppler sensing [7] are more advanced technologies to measure these vital signs. But, as subjects are exposed to some active emission in these methods, it is uncertain whether their long-term usages are safe. Recently, in pioneering work by Pavlidis and his colleagues [6, 9], a novel contact-free vital sign detection technique has been introduced and explored which is based on the infrared light emitted by the human body itself. This approach has the merits of low risk of harm and convenience for quick deployment. As heart and respiratory rates can be detected safely and wirelessly, it is more feasible to use these parameters in many applications, such as airport health screening, long-term elder care, and workplace preventive care.

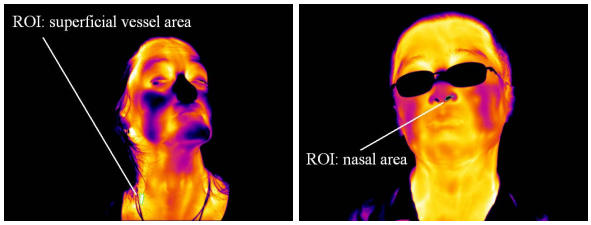

Figure 1. Illustration of regions of interest for heart and respiratory estimation from infrared video.

Periodic temperature changes of certain skin area can be detected by passive thermal video and reveal the associated heart or respiratory rates. The pulsating blood flow induces subtle periodic temperature changes to the skin on top of superficial vessels by heat diffusion [2]. The temperatures of inhaled and exhaled air are noticeably different. These temperature modulations can be detected through pixel intensity changes in certain regions of interest (ROI) using a high-sensitivity passive infrared camera. The corresponding heart or respiratory rates can be measured quantitatively by harmonic analysis of these changes. As illustrated in Fig. 1, the ROI for the measurement of temperature modulation is the skin area above the carotid artery for heart rate and the exhaled breath from the nares for respiratory rate.

In practice, heart or respiratory rate estimation through thermal video presents two fundamental challenges: accurate subject alignment for temporal signal extraction and robust harmonic analysis with low signal-to-noise ratio (SNR) temperature modulation signal. Harmonic analysis of temporal signal is only meaningful when accurate correspondences between pixels in consecutive infrared video frames

[1] The work was done during the internship of the first author in FXPAL.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

are available. Although our subjects attempted to keep still, involuntary muscular movements and slight voluntary motions are inevitable in a relatively long duration such as tens of seconds. Thus, in order to extract the temperature variation of an ROI pixel over time, it is critical to align the ROI by compensating the subject motion.

On the other hand, given the pixels are correctly associated, the harmonic analysis is still hard because of the low SNR and other distractions. The temperature modulation due to heart beat is usually less than 0.1K (Kelvin), which is far less than the normal human temperature of around 310K. Moreover, the temperature sensitivity of the state-of-the-art infrared camera we used is about 0.025K, which is close to the modulation signal level, resulting in a low SNR. Additionally, occlusions of the skin by objects such as clothing, hair or jewelry can also complicate the task since they have different heat diffusion properties from that of skin. These facts make it hard to identify correct heart beat signal through harmonic analysis, since the raw temperature of a pixel is usually noisy, as shown in Fig. 2.

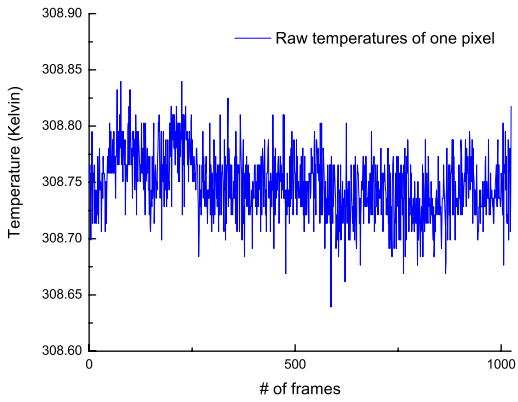

Figure 2. Raw temperature of a pixel on top of a superficial vessel.

Confronted by the aforementioned difficulties, existing methods resort to questionable assumptions to align subjects and simple heuristics to suppress noises. In [9], the vessels are located manually and regarded as long narrow structures. In addition, it is assumed that the measured subject is perfectly motionless. Further, [10, 7] align the vessel region using the hottest pixels to track the center of blood vessels. However, since the derivative around the maximal temperatures are relatively small, tracking based on those hottest spots tends to be unstable. [3, 4] attempt to align subjects based on foil makers placed on skin that are not available under most measurement scenarios. To combat noise, the signals of pixels in ROI are averaged in both spatial and frequency domain [9, 10, 7, 3, 4]. Moreover, there are a few ad hoc parameters that are difficult to incorporate into an automatic detection system, such as the width of the vessel [9], the sizes of the ROI [3], and the band width in the inverse wavelet transform [4].

Another natural and critical question that remains unclear in the existing methods is whether the temperature modulation does exist in the selected region of interest. As the temperature modulation level due to blood pulsating is far less than normal skin temperature, the subtle changes in the raw thermal video are unnoticeable to human operators. So, we design an effective way to visualize the subtle periodic changes qualitatively based on frame differences. After subject alignment, the frame differences against certain reference frame are calculated for every frame. An example is shown in Fig. 3, where the temperature differences of the ROI close to a superficial vessel exhibit visible periodic changes in video with 60Hz frame rate. The temperature differences look like random noise in the first several frames, but a definite pattern appears around 0.433 second at frame 26, close to one half of the pulse period. The temperature differences become noisy again around 0.983 second at frame 59. This observation confirms that there does exist periodic temperature fluctuation and implies that all pixels are not equally informative. Thus, averaging the signals of different pixels may obscure the modulation signal. Note, this visualization method can also be used to verify the subject alignment performance subjectively.

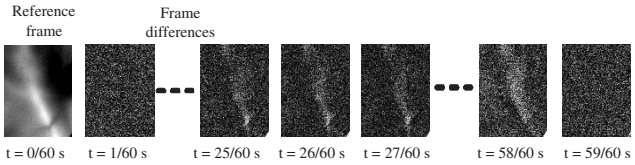

Figure 3. Visualization of the periodic changes of temperature differences in a thermal video with frame rate 60Hz.

In this paper, we propose a novel approach to heart rate estimation by harmonic analysis of the skin temperature modulation around superficial vessels. The method can be also applied to respiration rate estimation by measuring the periodic temperature changes around nasal area. The new approach incorporates three modules: subject alignment by motion compensation, signal enhancement and denoising, and robust harmonic analysis. First, we propose an automatic contour segmentation and tracking algorithm for tracking a region near a superficial blood vessel to align every pixel frame by frame, without requiring markers or predefined parameters. Then, to enhance the signal, we apply a fast Fourier transform (FFT) to the temperature signal of each pixel over time in a sliding window, followed by a non-linear filter to reduce the low-frequency signal leakage in the measurement frequency range. The observation that some pixels may be less informative than others motivates us to leverage a clustering procedure to remove the

outliers. Afterwards, the pixels in the selected cluster vote for the dominant frequency component and the final estimated heart rate is post-processed by a median filter.

The proposed method requires less human intervention and is robust to gentle subject movement, facial expressions, and noise, so that it exhibits more stable estimation results than the previous method [9, 7]. Evaluated on thermal video sequences of 20 subjects captured by a mid-wave infrared camera, the average differences of the estimated heart rates from the ground truth range from 0.2 to 3.4 bpm with less than 4.9 root mean square error (RMSE). To the best knowledge of the authors, we first show the heart rate estimation with respect to time and the variance of the estimates, which demonstrate the stability and reliability of the proposed approach. The same method also achieves promising results on respiration rate estimation.

The related work is discussed in Sec. 2 and the overview of our approach is presented in Sec. 3. Then, three primary modules, *i.e.* subject alignment, signal enhancement, and harmonic analysis, are described in detail in Sec. 3, Sec. 4, and Sec. 5, respectively. The experimental evaluation is given in Sec. 7 with the concluding remarks in Sec.8.

2. Related work

Recently, [9] proposed to estimate average cardiac heart rates from thermal imaging of major superficial vessels on faces and demonstrated preliminary results. The locations of the vessels were manually labelled and it was assumed that the measured subject was perfectly motionless. The major blood vessels were assumed to be long narrow structures. A fixed number of pixels on the normal line directions were averaged to suppress noise. The FFT spectrum of the temperature of each pixel was averaged with the historical spectrums and the dominant frequency in the band of 40-100 beats per minute (bpm) was regarded as the heart rate. The method was further improved in [10, 7] to employ the hottest pixels to track the center of blood vessels in order to align the vessel regions. Since hottest pixels may not exactly correspond to the center of the vessels and the hot regions may be flat, tracking based on those hottest spots tends to be noisy and unreliable.

[3, 4] described a Superficial Temporal Artery (STA) measurement model to estimate heart rates based on arterial wall volumetric change corresponding to blood pressure modulation. Subjects were aligned based on foil makers placed on their skin. The ROI close to superficial vessels was divided into multiple cells and averaged. Then, one cell was picked to estimate the transient heart rate and waveform using band-pass filtering. The criteria to select ROI scale, the optimal ROI cell, and the band width in the inverse wavelet transform [3, 4] were not fully justified.

In contrast to previous methods, other than giving a rough initial region, our approach makes no assumptions about the location and scale of the vessel used in estimation. The regions of interest for analysis are automatically segmented and tracked, so that they are robust to gentle subject motion and facial expressions and applicable in cases where the skin area is partially occluded by hair or jewelry. In addition, as we employ a clustering procedure to remove some outliers, the results are more stable than those obtained by directly averaging the signals of all pixels in ROI.

3. Overview of our approach

The proposed approach is composed of three building blocks: subject alignment by contour segmentation and tracking, signal enhancement with a non-linear filter and outlier removal by a K-means clustering, and dominant frequency component by voting. The entire procedure is summarized in Fig. 4.

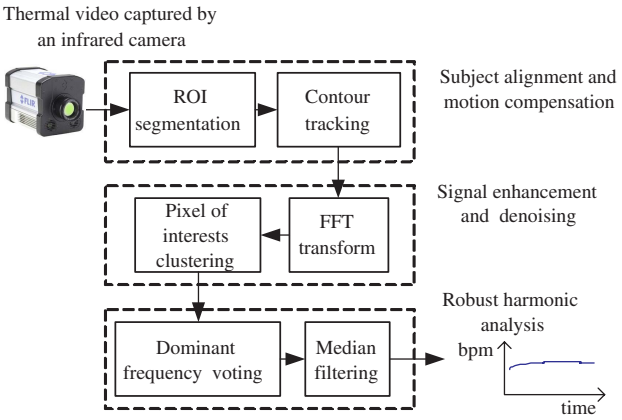

Figure 4. The block diagram of our approach.

Given the thermal video sequence, we define the region of interest as the skin area on top of a superficial vessel. First, a large bounding box is manually cropped as a rough initialization. Then, we compute all the *isotherms* with a small temperature step. This incremental temperature step is determined based on the temperature modulation level and the sensitivity of the camera. The contour corresponding to the steepest temperature changes specifies the boundary of the ROI. This contour is least sensitive to small temperature variations and capable of addressing the question of whether the temperature changes are due to subject motion or blood pulsating.

This contour enables it to mimic artificial markers, moreover, since the human body is not rigid, the contour inside the measurement region is less affected by movement than artificial markers placed farther away from the critical measurement region. Since the ROI inside the contour is very small, it is reasonable to assume most if not all pixels belong to one rigid object. Therefore, after contour tracking

using an active contour model [8], we can establish the correspondences of pixles between frames and align the pixels within the contour based on the center and size of the ROI. In this way, the temperature variation can be extracted for each pixel for further harmonic analysis.

After applying an FFT on the windowed raw temperature signal of each pixel, we employ a non-linear filter to reduce low frequency leakage. As can be seen in Fig. 3, all pixels do not contain the same level of periodic changes, and their phases could be different due to heat diffusion. Thus, rather than averaging the signal in either the time or frequency domain, we cluster the pixels according to the frequency components in the band of interest and discard the minor clusters as outliers. The clustering procedure largely reduces the influence of noise and other small temperature distractions. The largest cluster is selected to vote for the dominant component frequency. This is post-processed by a median filter to produce the final estimate.

4. Subject alignment and motion compensation

Due to heat diffusion [2], the skin temperature on top of the superficial vessels are generally higher than that of the nearby skin. We utilize this property to select and align the ROI for subjects. This is done by selecting an optimal isotherm that corresponds to the sharpest spatial temperature change and tracking this contour to compensate for an ROI's motion. Specifically, the captured infrared video sequence is denoted as $\{I_0, I_1, \cdots, I_t, \cdots\}$ where the temperature of one pixel $\mathbf{x} = (u, v)$ at frame t is denoted by $I_t(\mathbf{x})$. We first segment the ROI R_0^* from I_0 which is associated with the skin area on top of a superficial vessel, then track the ROI R_t^* in successive frames to compensate for any subject movement.

4.1. Automatic ROI segmentation

For the first frame I_0, the rough initialization region is denoted by R^0. Given a temperature step ΔI and a reference temperature I_{ref} that is the average temperature of the pixels on the bounding box of R^0, we segment the initialization region R^0 into a series of nested regions, $R^i \supset R^{i+1}$. The boundary of R^i, *i.e.* the isotherm, is defined by $\Gamma^i = \partial R^i$ with the property that for each pixel $\mathbf{x} \in \Gamma^i$, there exists at least one pixel in its neighborhood $N(\mathbf{x})$ whose temperature is higher than $I_{ref} + i\Delta I$, *i.e.*

$$\forall \mathbf{x} \in \Gamma^i, \exists \mathbf{x}^{'} \in N(\mathbf{x}), \text{ s.t. } I_0(\mathbf{x}^{'}) > I_{ref} + i\Delta I, \quad (1)$$

where $N(\mathbf{x})$ is the 4-connected neighborhood of pixel $\mathbf{x}$. The incremental temperature step ΔI is chosen as 0.1K based on the temperature modulation level (0.08K) and the sensitivity of the camera (0.025K). More specifically, if the incremental step is smaller than 0.08K, not only determined by the subject motion solely, temperature changes due to blood pulsating could also affect the ROI segmentation and tracking. On the other hand, if the step is too large, there could be large changes in the derived regions of interest across successive frames.

The optimal threshold, $I^* = (I_{ref} + i^*\Delta I)$, is selected such that the area of the region R^{i^*} has the steepest drop against the area of R^{i^*-1}, so R^{i^*} is less sensitive to small temperature fluctuations. The largest connected component in R^{i^*} is selected as the ROI R_0^* with its boundary Γ_0^* at the first frame I_0. These series of isotherms are efficiently calculated by a flooding procedure implemented by dynamic programming. Note that this is different from pure thresholding, as the temperatures of pixels inside R^{i^*} could be less than the optimal threshold I^*. This procedure is illustrated in Fig. 5.

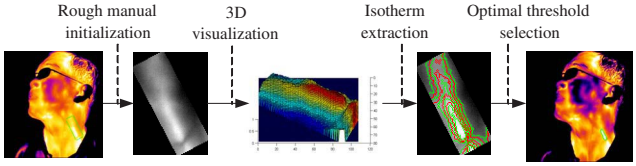

Figure 5. The procedure for automatic ROI segmentation.

4.2. ROI alignment by contour tracking

For the following frames, we employ the same threshold I^* to extract the isotherm and track the contour Γ_t^* by minimizing the energy similar to [8]. The energy of a contour Γ_t is defined as

$$\begin{aligned} E(\Gamma_t) &= \oint_{\Gamma_t} (E_{image}(\Gamma_t) + E_{ext}(\Gamma_t)) \quad (2) \\ &= \oint_{\mathbf{x}_j \in \Gamma_t} ||I_t(\mathbf{x}_j) - I^*||^2 + \alpha ||\mathbf{x}_j - \tilde{\mathbf{x}}_j||^2, \end{aligned}$$

where $\tilde{\mathbf{x}}_j$ is the closest point to $\mathbf{x}_j$ on the tracked contour in I_{t-1}, that is, $\tilde{\mathbf{x}}_j = \text{argmin}_{\mathbf{x} \in \Gamma_{t-1}^*} ||\mathbf{x} - \mathbf{x}_j||^2$, and $\alpha = 0.001$ is a coefficient to control the trade-off between consistency with the threshold I^* and temporal smoothness. The image energy E_{image} is the integration of temperature differences against the optimal threshold I^* selected at I_0. Unlike [8], we do not impose an internal smoothness constraint on the contour, but use a temporal smoothness constraint modelled by the external energy E_{ext}. The contour tracking result Γ_t^* minimizes the energy in Eq. 2 by a local gradient decent search.

After contour tracking, we align the pixels within the ROI R_t^* based on its gravity center and size. Thus, the temporal signals for individual pixels inside the ROI are extracted and denoted by $s(\mathbf{x}_j, t) = \{\cdots, I_{t-1}(\tilde{\mathbf{x}}_j), I_t(\mathbf{x}_j)\}$ for every $\mathbf{x}_j \in R_t^*$.

5. Signal enhancement and outlier removal

Since the temperature modulation magnitude 0.08K is approximately 4000 times less than the skin temperature 310K and only 3 times larger than the camera sensitivity, it is critical to enhance the signal and reduce the influence of noise before doing a harmonic analysis.

5.1. Non-linear band-pass filtering

For every frame, we analyze the temperatures signal $s(\mathbf{x}_j, t)$ of all pixels in R_t^* using a sliding window with N frames to estimate the underlying heart rates. We compute an N-point FFT of $s(\mathbf{x}_j, t)$ with a window function $W(t)$.

$$H(\mathbf{x}_j, f) = \mathcal{F}[W(t)s(\mathbf{x}_j, t)]. \quad (3)$$

Since the low frequency component of the thermal signal is several thousand times higher than the temperature modulation level caused by periodic pulsating blood flow, the low-frequency signal leakage due to a finite sliding window may overwhelm the modulation. A simple way to reduce disturbances from low frequency is a Hamming window, which has much smaller side lobes than does a rectangular window. However, to reduce nearby nuisance disturbances and increase frequency estimation resolution, we prefer the main lobe of the frequency response narrower than that of the Hamming window. Increasing the length of the sampling window can achieve that, but at the cost of longer measurement times. Instead we design a non-linear filter to tackle this low-frequency leakage problem by combining the advantages of rectangle and Hamming windows. After performing an N-point FFT on the windowed signal using both a rectangle window $W_r(t)$ and a Hamming window $W_h(t)$, we take the point-by-point minimum of these two spectrums in the frequency domain,

$$H(\mathbf{x}_j, f) = \min(\mathcal{F}[W_r(t)s(\mathbf{x}_j, t)], \mathcal{F}[W_h(t)s(\mathbf{x}_j, t)]). \quad (4)$$

Thus, the combined non-linear filter has as narrow a main lobe as a rectangular window yet efficiently reduces the signal leakage from low frequencies as would be expected by a Hamming window. Fig. 6 shows the frequency responses of a rectangular and a Hamming window for one sinusoid signal, contrasted with the frequency response of the combined filter drawn with red line.

5.2. Pixels of interest clustering

The temporal signal $s(\mathbf{x}_j, t)$ extracted after ROI alignment is vulnerable to noise and short term disturbances. In the existing approaches, signals of the pixels in the ROI are averaged in either the spatial domain or the frequency domain to reduce the influence of noise. However, as shown in Fig. 3, the pixels may have unknown phase shifts and some may not have clear temperature modulation. Therefore, in contrast to other approaches, we propose to use a clustering procedure to remove outliers, with the result that the ROI is more robust to short time disturbances in a small region.

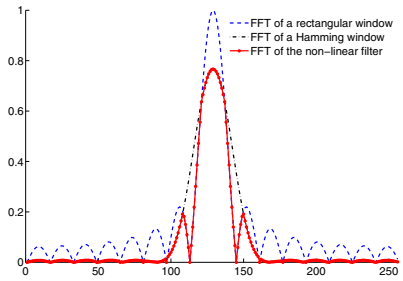

Figure 6. Illustration of the frequency response of the non-linear filter with a 256-point FFT.

The frequency components of $H(\mathbf{x}_j, f)$ in the band of interest (40-100 bpm for heart rates, and 6-30 bpm for respiratory rates) are represented as an M-dimensional feature vector $\{h_1(\mathbf{x}_j, f), \cdots, h_M(\mathbf{x}_j, f)\}$. Then, these features for all pixels in R_t^* are feed into a K-means clustering module. The cluster number K is set as the number of frequency components M empirically. After the clustering, the largest cluster is selected to estimate heart beat rates. We refer to the set of pixels in this cluster as the pixels of interest. This is based on the assumption that the outliers are sparsely distributed. If more prior knowledge about the frequency response is available, *e.g.* the prior distributions of heart or respiratory rates, we can use a different criterion to select the cluster. Though an individual signal $s(\mathbf{x}_j, t)$ can be very noisy, we can isolate the disturbances by performing the cluster analysis, thus effectively reducing their impacts on the final estimation.

6. Robust harmonic analysis

We determine the dominant frequency in the band of interest by majority voting, as the mean spectrum [9, 7] tends to be vulnerable to some abrupt distractions, such as involuntary facial expressions and partial occlusions by hair. The frequency peaks of $\{h_1(\mathbf{x}_j, f), \cdots, h_M(\mathbf{x}_j, f)\}$ in the set of pixels of interest are used to vote for the dominant frequency component. The bin with the most votes is selected as the dominant frequency, as shown in Fig. 7. The final heart rate estimates are the median filtered results of the dominant frequency components in a small sliding window (1 second in the experiments below).

7. Experimental results

7.1. Experiment settings

In the experiments, we use a mid-wave infrared camera [5] that can capture infrared light in the range of 3.0-5.0

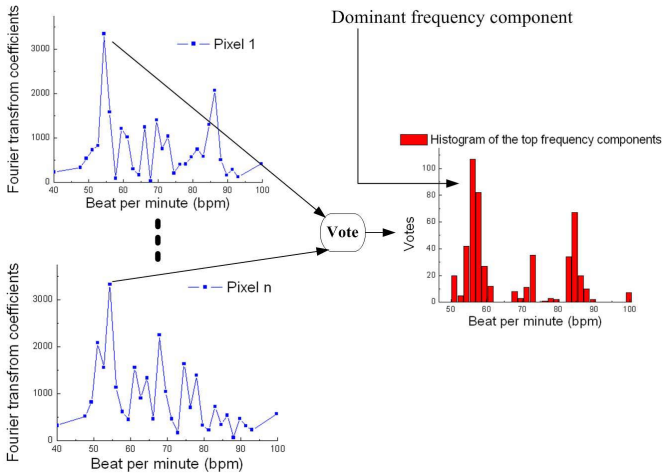

Figure 7. Voting for the dominant component.

microns with the temperature sensitivity 0.025K. Twenty people including 6 females and 14 males between the ages of 20 to 60 volunteered to participate. For each subject, we captured 5 one-minute thermal video clips at a resolution of 640×512 with 14 bits for one pixel. The testing sequences had three different frame rates 30, 60 and 115 frame per second (fps) with the corresponding sliding window length $N = 1024, 2048$, and 4096 respectively. The band of interest has $M = 34$ components. The latency of the estimation is about 34 seconds.

7.2. Ground truth acquisition

During the capture of the thermal video, the ground truths (GT) for subjects' heart rates were detected by a piezoelectric pulse transducer and ECG electrodes, while the ground truths for respiratory rates were measured by a piezoelectric respiration transducer. These ground truths were recorded by a PowerLab data acquisition system [1] and synchronized with the thermal video using time stamps. One thousand data points were sampled and recorded per second. The entire setup is shown in Fig. 8.

To obtain the ground truth rates from the waveforms (Fig. 9 shows some representative waveforms), we averaged the intervals between the local maxima in a small time window to estimate the GT periods. The length of this window corresponded to the highest possible rate. Specifically, we employed windows of 0.333 seconds (180 bpm) for heart rate estimation and 2 seconds (30 bpm) for respiratory rate estimation. After that, the GT rates obtained by the piezoelectric pulse transducer and ECG electrodes were averaged to generate the GT heart rates used in the evaluation.

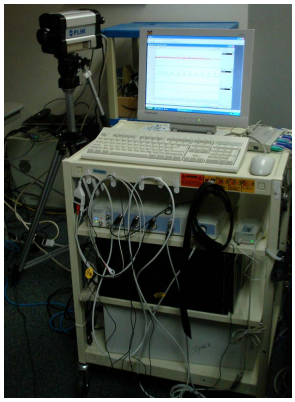

Figure 8. Illustration of the experiment setup.

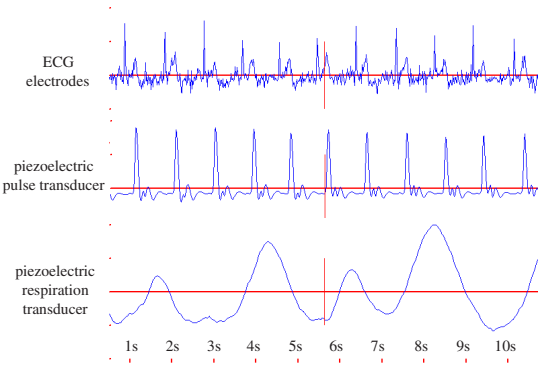

Figure 9. Illustration of the ground truth waveforms.

7.3. Heart rate estimation

We compared the proposed approach with the state-of-the-art heart rate estimation algorithm [9, 7]. The window size used in [9, 7] is 512 points at frame rate 30 fps. That window size limits the frequency resolution to 0.059Hz, or 3.52 bpm. Since 3.52 bpm is usually greater than people's heart beat variation in one minute, we increased the window size to reduce possible aliasing. For video sequences at frame rates 30, 60, and 115 fps, we used 1024, 2048, and 4096 points, respectively, as N in the FFT. Our band of interest was 40-100 bpm, the same as [9, 7], with $M = 34$ for all sequences. Fig. 10 illustrates the initial ROI segmentation results drawn as green contours for all subjects.

Table. 1 shows the ground truth heart rate, and the estimated average heart rate (Est. bpm), average difference (Diff. bpm), and RMSE against the GT rate (GT bpm) for our method and for the comparison method (indicated by a postfix *(S)*), where possible. Note, since [9, 7] require that subjects are completely still and assume that blood vessels are long narrow structures, it is not applicable for the majority of our test sequences. For 4 sequences where the baseline system [9, 7] can yield reasonable results, the proposed method consistently outperforms [9, 7], both in terms of average difference in bpm and RMSE. Two representa-

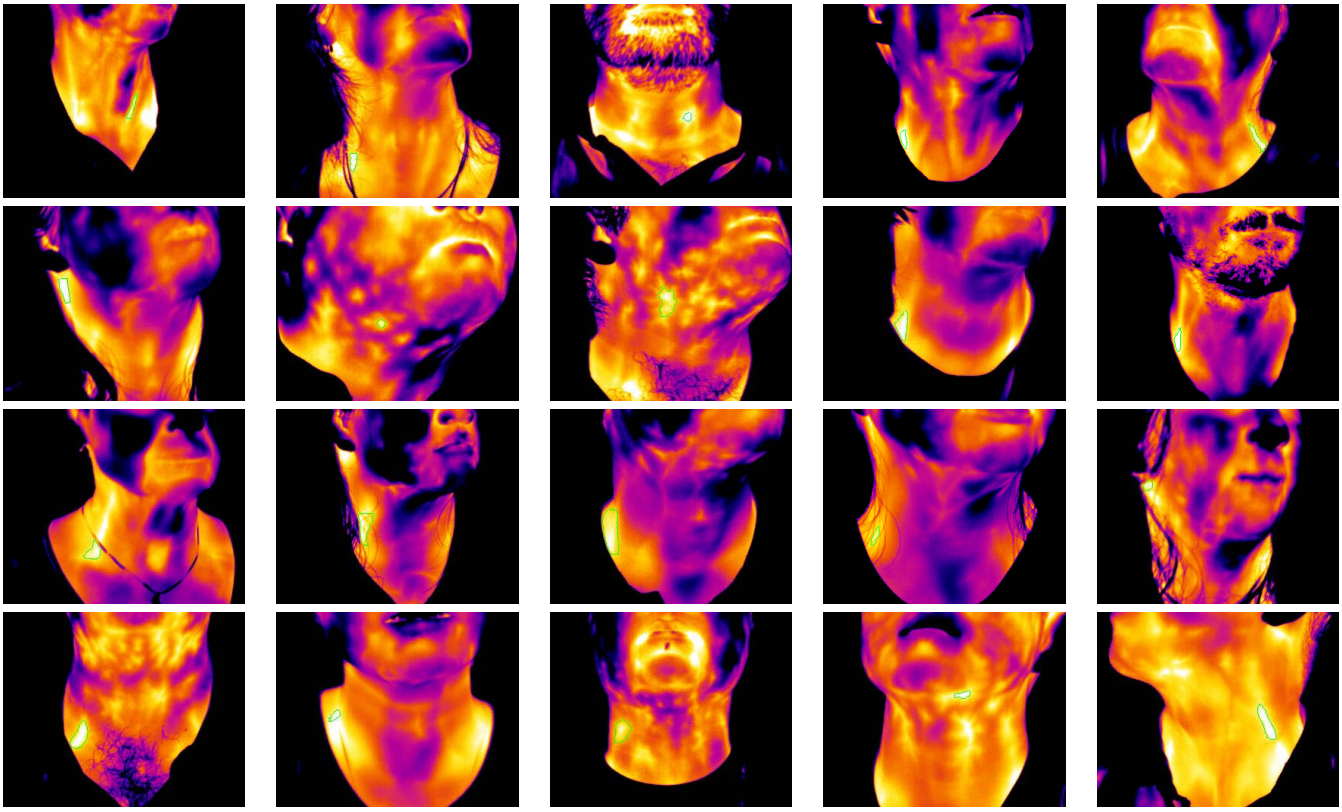

Figure 10. The initial ROI segmentation results of all subjects which are drawn as green contours.

tive point-by-point comparisons of both methods with the GT is shown in Fig. 11. It shows that the proposed method exhibits more stable estimation than [9], largely due to our more accurate ROI alignment and the removal of outliers representing significant noise in the pixels. Note that the 34-second latent periods are omitted in Fig. 11.

7.4. Respiratory rate estimation

The proposed approach was also applied to respiratory rate estimation with the same parameters except that the isotherms in descending order in ROI segmentation, *i.e* $I_0(\mathbf{x}^{'}) < I_{ref} - i\Delta I$ in Eq. 1. The frequency band of interest was set as 6 to 30 bpm. The comparison with the GT rates is listed in Table. 2. The average differences are less than 2.2 bpm. Because the nasal area of the subject had to be clearly visible, only test video clips for 8 subjects yielded reasonable estimates.

Table 2. Respiratory rate estimation results.

Subject #	fps	# of frames	GT bpm	Est. bpm	Diff.
4	60	3000	18	15.8	-2.2
7	60	3000	17	15.1	-1.9
10	105	5000	11	11.8	+0.8
11	105	5000	17	16.8	-0.2
14	105	5000	16	13.9	-2.1
15	105	5000	15	13.1	-1.9
17	105	5000	20	18.5	-1.5
19	105	5000	16	15.2	-0.8

8. Conclusion

In this paper, we propose a new vital sign measurement approach using passive thermal video. Specifically, heart and respiratory rates are estimated by extracting the dominant frequency component of the periodic temperature changes around the skin area on top of a superficial vessel or the nasal area. Human intervention is minimized by automatic contour segmentation and tracking. The signal, which has a low SNR, is enhanced by non-linear filtering and clustering of pixels of interest. Consequently, the proposed method is insensitive to initialization and robust to gentle subject movement and facial expressions. The experiments demonstrate more stable estimation compared with the state-of-the-art method and show that this new method is promising for quick vital sign measurement in unconstrained environments.

References

[1] ADInstruments. http://www.adinstruments.com. 6

Table 1. Heart rate estimation results.

Subject #	fps	# of frames	GT bpm	Est. bpm	Diff. bpm	RMSE	Est. bpm (S)	Diff. bpm (S)	RMSE (S)
1	30	2000	65.3	65.8	+0.5	1.9			
2	30	2000	66.6	63.9	-2.7	3.9			
3	30	1750	65.7	64.7	-1.0	3.3			
4	60	3000	59.8	60.7	+0.9	2.5	61.4	+1.6	3.3
5	60	3500	60.7	60.3	-0.4	3.3	56.1	-4.6	8.2
6	60	2500	66.3	63.0	-3.3	3.9			
7	60	3000	61.1	60.9	-0.2	2.3			
8	115	5000	64.0	65.0	+1.0	3.8			
9	115	5000	78.9	80.1	+1.2	1.9			
10	115	5000	65.2	64.4	-0.8	1.7			
11	115	5000	62.8	66.2	+3.4	4.2			
12	115	5000	63.5	62.4	-1.1	3.1			
13	115	5000	73.3	72.6	-0.7	1.8			
14	115	5000	86.6	87.9	+1.3	4.9	88.9	+2.3	5.8
15	115	5000	78.7	76.5	-2.2	3.1			
16	115	5000	75.3	74.7	-0.7	1.9			
17	115	5000	83.1	83.2	+0.1	2.1			
18	115	5000	67.2	68.2	-1.0	1.3			
19	115	5000	67.6	69.3	+1.7	2.8	64.6	-2.0	6.4
20	115	5000	68.7	70.1	+1.4	2.9			

[2] H. Arkin, L. X. Xu, and K. R. Holmes. Recent developments in modeling heat transfer in blood perfused tissues. *IEEE Trans. Biomed. Eng.*, 41(2):97 – 107, Feb. 1994. 1, 4

[3] S. Y. Chekmenev, A. A. Farag, and E. A. Essock. Multiresolution approach for non-contact measurements of arterial pulse using thermal imaging. In *CVPR'06 Workshop*, page 129, NYC, NY, June 17 - 22, 2006. 2, 3

[4] S. Y. Chekmenev, A. A. Farag, and E. A. Essock. Thermal imaing of the superficial temporal artery: An arterial pulse recovery model. In *IEEE Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (OTCBVS'07)*, pages 1 – 6, Minneapolis, MN, June 22, 2007. 2, 3

[5] FLIR Systems. www.flirthermography.com/cameras /camera/1086/. 5

[6] M. Garbey, A. Merla, and I. Pavlidis. Estimation of blood flow speed and vessel location from thermal video. In *CVPR'04*, volume 1, pages 356 –363, Washington, DC, Jun.27 - Jul.2, 2004. 1

[7] M. Garbey, N. Sun, A. Merla, and I. Pavlidis. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Trans. Biomed. Eng.*, 54(8):1418 – 1426, Aug. 2007. 1, 2, 3, 5, 6

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Journal of Computer Vision*, 1(4):321 – 331, Jan. 1988. 4

[9] N. Sun, M. Garbey, A. Merla, and I. Pavlidis. Imaging the cardiovascular pulse. In *CVPR'05*, volume 2, pages 416 – 421, San Diego, CA, June 20 - 25, 2005. 1, 2, 3, 5, 6, 7

[10] N. Sun, I. Pavlidis, M. Garbey, and J. Fei. Harvesting the thermal cardiac pulse signal. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06)*, volume 2, pages 569 – 576, Copenhagen, Denmark, Oct. 1 - 6, 2006. 2, 3

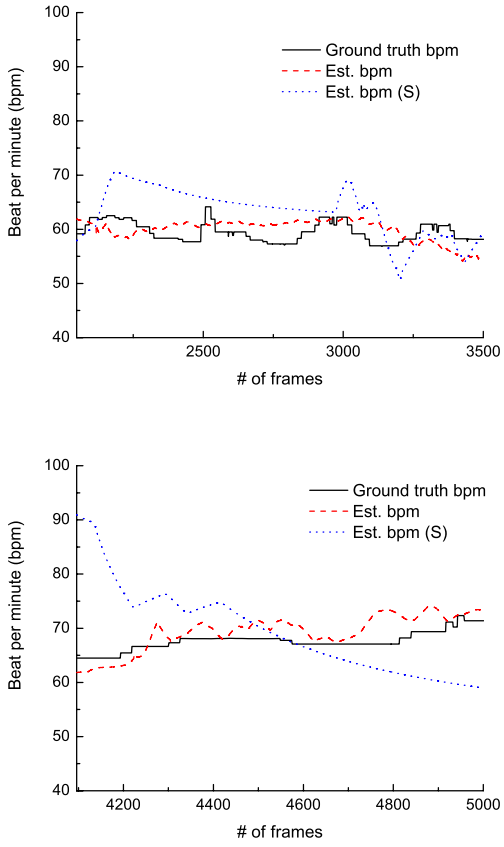

Figure 11. Point-by-point comparison of estimated heart rate with the ground truth for subject #4 at 60 fps (top) and subject #19 at 115 fps (bottom).