Non-contact, Wavelet-based Measurement of Vital Signs using Thermal Imaging

S. Yu. Chekmenev, H. Rara, and Aly A. Farag

Computer Vision and Image Processing Laboratory
University of Louisville, Louisville, KY 40292, USA
{sergeic, ham, farag}@cvip.uofl.edu
http://www.cvip.louisville.edu

Abstract

In this paper we evaluate the potential of using non intrusive remote passive thermal imaging for measurements of human vital signs. Breathing causes noticeable changes in temperature at the nasal area, which appear as periodic changes in the face thermogram. The major arteries of the body produce time-varying heat patterns which yield information about the cardiac cycle. We focus on thermal imaging using long-wave IR of the face and neck areas. A new method based on wavelet analysis is developed to extract the pulse and respiration. Experimental results on four human subjects are provided, and validated against standard approaches for measuring heart rate and respiration. The proposed method has potential applications in non-contact vital signs monitoring and intent identification at a distance.

Keywords: Wavelets, Thermogram images, physiological variables, intent identification.

1. Introduction

Vital signs are physical signals that indicate a living person, which include heart rate, breathing rate, temperature and blood pressure. The signs can be measured and used to assess the person's level of physical functioning. Normal ranges of vital signs vary with age, sex, weight, exercise tolerance and body conditions [1]. In this paper, we focus on the measurement of the breathing rate and heart rate. The baseline standard for pulse measurement is the electrocardiogram (ECG). Aside from measuring the rate and regularity of heartbeats, it is used also to detect any presence of damage in the heart. The ECG is a contact type of measurement which the emitted energy from the objects may provide great promise in distant measurements of the vital signs without side effects. Pavlidis, et al. [4, 8] used mid-wave IR sensors for the measurement of cardiac and breathing rates at a distance. The cardiac pulse measurement is based on the findings in [4] that the skin temperature in the vicinity of the major arteries is directly related to the pulse waveform but the exact shape is smoothed, shifted and noisy with respect to the actual pulse because of the diffusion process and air flow. To solve for the heart rate, Pavlidis used the Fourier Transform (FT) to isolate the relevant frequency components related to the heart signal.

Holdsworth [5] investigated the blood-velocity waveforms in the left and right common carotid arteries using a pulsed-Doppler ultrasound with ECG data as the baseline measurement. Geisheimer [6] developed a Radar Vital Signs Monitor (RVSM) that uses an active radar detector in which the reflected wave provides information about the motion of the chest and body due to the cardiac and respiration cycles. The Doppler and Radar devices have the disadvantage of being active sensors, that is, the energy that is focused may have harmful side effects on the health of the individual [7]. In contrast to active sensors, passive infrared (IR) detectors which measure the emitted energy from the objects may provide great promise in distant measurements of the vital signs without side effects. Pavlidis, et al. [4, 8] used mid-wave IR sensors for the measurement of cardiac and breathing rates at a distance. The cardiac pulse measurement is based on the findings in [4] that the skin temperature in the vicinity of the major arteries is directly related to the pulse waveform but the exact shape is smoothed, shifted and noisy with respect to the actual pulse because of the diffusion process and air flow. To solve for the heart rate, Pavlidis used the Fourier Transform (FT) to isolate the relevant frequency components related to the heart signal.

Our approach is most related to that of Pavlidis and his group. However, we show that long-wave IR can indeed provide an excellent waveform of the heart signal which can easily provide the heart rate. For breathing measurements, Pavlidis used the air exhaled, to measure the breathing rate. The consequence of this is that the subject must have a side-view orientation to the camera [8]. Our approach focuses on the changing temperature in the nasal region of the thermal image in both front and profile views. Pavlidis uses the raw thermal data directly to construct time-varying signals for each of the pixel from area of interest. This makes them extremely noisy. Averaging of these signals may not improve the signal quality as desired. In our approach we first represent thermal the image at different scales. Then we choose the scale that carries the most information relevant to breathing and heart signals, and provides more representable features of the region of interest (ROI) than
other scales. Only after these steps, we construct time-varying signals from each point in the ROI from this new representation. Averaging and continuous wavelet analysis of these waveforms have been shown to yield excellent result.

The remainder of the paper is organized as follows: Section (2) focuses on the basics of thermal imaging, breathing function and pulse physiology, and the wavelet theory. Section (3) emphasizes on the experimental setup.

2. Theory

Thermography makes use of the infrared (IR) spectral band of the electromagnetic spectrum. IR involves four bands: near infrared (0.75-3 µm), middle infrared (3-6 µm), far infrared (6-15 µm) and extreme infrared (15-100 µm). Infrared video cameras are passive (emits no energy), but merely collects the thermal radiation emitted from the surface of the human body [9]. The IR camera used in this paper operates in the far infrared range.

The breathing cycle is composed of three parts: inhalation, exhalation, and post-exhalation pause [8, 10]. In this paper, we consider the post-exhalation pause to be a short instant at the end of the exhalation phase, thus, the breathing function may be considered as a two-part process of inhalation and exhalation.

The heart rate is the number of contractions of the heart in one minute and is measured in beats per minute (bpm) [11]. The pulse is the straightforward way of measuring the heart rate. Except in the case of arrhythmias, the heart rate and pulse are quite similar. There are various points in the body where the pulse signals can be felt. In this paper, we focus on the carotid vessel complex in the neck and the superficial temporal vessel complex because they represent the most accessible parts in the body for the IR camera. The carotid vessel complex is composed mainly of the jugular vein and the various carotid arteries in the area. The carotid arteries, which give the strongest pulse signal, unfortunately, are mostly under the muscles (sternocleidomastoideus) of the neck. The presence of superficial vessels (veins and capillaries) in the area, together with motion artifacts, makes the extraction of the pulse signal at a distance a challenging proposition.

Multi-Scale Image Decomposition (MSD): Multi-resolution analysis was formulated by Mallat [13]. A large body of research exists on using wavelets for signal and image analysis (e.g., [13]-[15]). In this paper, we use a pyramid structure [15] for implementing MSD of thermal images. We use kernels obtained from the second derivative of the Gaussian function known as the Mexican hat (MH) function (see Figure 2):

\[
MH(t) = (2 - t) \cdot \exp\left(-\frac{t}{2}\right), \quad t = \left(\frac{x-b_x}{a} \right)^2 + \left(\frac{y-b_y}{a} \right)^2
\]

where \(b(b_x, b_y)\) is a translation vector and \(a\) is scalar. The MH kernels are orthogonal to all their dilations and shifts, have zero average value and compactly supported. These wavelets enable us to effectively represent thermal images at different scales and locations. Figure 2 shows the thermal image decomposition using MH kernels at three different scales. From the figure, it is clearly seen how the coarse-to-fine representation automatically locate heat flashes on the upper head and neck areas where the cardiac pulse can be measured.

The carotid arteries, where the pulse can be extracted, are mostly under the muscles of the neck. We chose the area of thermal images that provide best information about the pulse.

Figure 1: The carotid vessel complex [12].

Figure 2: Three-Scale MSD using MH wavelet kernel

The formal experimental procedure is composed of the following steps:

1. A human subject is seated stationary one meter away from the IR camera, which is focused on the face.
2. We acquire \(N\)-frames from the face and neck areas.
3. The carotid artery area of the neck is manually selected as the Region of Interest (ROI-1), as shown in Figure 3(b).
4. Manually select the nasal region as the second ROI (ROI 2), as illustrated in Figure 3(a).
(5) Compute the three-scale decomposition of each of the \(N\)-frames of thermal images.

(6) Compute the mean value of ROI-1 and ROI-2 for each scale for all \(N\) frames, and plot the average value with respect to time (frame), resulting to three 1-D plots.

(7) Apply continuous wavelet analysis on the resulting 1-D plots.

**Continuous Wavelet Analysis:** The continuous wavelet transform (CWT) of the function \(f(t)\), is defined as [17]:

\[
\psi(t) = \frac{1}{\sqrt{C_\psi}} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) f(t) dt,
\]

\[
C_\psi = \int_{-\infty}^{\infty} \left| \hat{\psi}(\omega) \right|^2 d\omega < \infty
\]

where the constant \(C_\psi\) in Eq. (3) is defined by the Fourier transform of the function \(\psi(t)\):

\[
\hat{\psi}(\omega) = \int_{-\infty}^{\infty} \psi(t) e^{-i\omega t} dt
\]

Eq. (3) implies that \(\psi(\omega) = 0\), when \(\omega = 0\). The function \(\psi(t)\) is called the mother wavelet. By shifting in time and dilating or compressing this function in the frequency domain, one obtains a set of self-similar functions, \(\psi_{a,b}(t) = a f(t) \psi\left(\frac{t-b}{a}\right)\), where \(a = f^{-1}\) is the scale that provides dilating or compressing, \(b\) is the time shift, and \(t\) is "time", which can also be frame, mass, energy etc., depending on the problem we deal with. The self-similar functions \(\psi(t)\) have to satisfy the following conditions: (i) \(\int_{-\infty}^{\infty} \psi(t) dt = 0\), \(\psi(t)\) are oscillating functions bounded at the origin; (ii) \(\int_{-\infty}^{\infty} \psi_k(t) \psi_l^*(t) = \delta_{kl}\), \(\psi(t)\) represents an orthonormal basis. The indexes \(k, l\) specify particular functions from the orthonormal basis (set of functions ordered by index). If the integrals (3) and (4) are bounded, then there exists inverse continuous transform (ICWT) which gives us the reconstructed signal:

\[
f(t) = \frac{1}{\sqrt{C_\psi}} a^2 \int_{-\infty}^{\infty} w(a, b) \psi\left(\frac{t-b}{a}\right) db ds
\]

In the equation above, we add the average value of a signal to the inverse wavelet transform in order to obtain a reconstructed signal. We do this because the average value of any wavelet is zero. In this study, we use the “Mexican Hat” (MH) wavelet, which is characterized by its good localization in the "time domain" and small number of oscillations:

\[
\psi(t) = (1 - t^2) e^{-t^2/2}
\]

The MH wavelet isolates local minima and maxima of the signal at the selected duration and allows us to analyze individual event localization more precisely.

---

*Figure 3: Manually selected ROI in the Carotid Vessel Complex off the center of the jugular vein.*

Before applying this approach to real signals lets see how it works for analytical one. Figure 4 (a) shows a plot of the sum of Lorenz-like functions defined as:

\[
f(t) = \frac{30}{(t-30)^3 + 50} + \frac{35}{(t-80)^3 + 60} + \frac{70}{(t-120)^3 + 120} + \frac{35}{(t-160)^3 + 60} + \frac{30}{(t-200)^3 + 55} + \frac{30}{(t-260)^3 + 50}
\]

In Figure 4(b), a Gaussian zero-mean noise is added to it. Figure 4 (c-d) shows the CWT of analytic signal with and without the noise, respectively. The application of a CWT to a one-dimensional function results into a two-dimensional function. Therefore, the CWT plane provides a powerful tool for analysis since it is possible to see the signal content at different scales at different time instances. At fine scales, we observe the Gaussian noise, which limited within a narrow band and do not continue towards the coarse or bigger scales. As we go up in CWT plane, Figure 4(c-d), we begin to observe structures related to the Lorenz-like functions. Finally, at the highest scales, we see the coarsest components of the analytic signal that span substantial amount of scales and time interval.

---

*Figure 4: (a) Nasal Area (ROI-2) Segment of Carotid Vessel Complex (b) Manually selected ROI-1 on the Segment of Carotid Vessel Complex (c) ROI-1 at Scale 1 decomposition (d) ROI-1 at Scale 2 decomposition (e) ROI-1 at Scale 3 decomposition*
Knowing which scales the important structures belong to, the analysis of them can be done very efficiently. The components that are beyond the scale-band of our interest can be simply disregarded. Those which are inside the band can be removed by the threshold, since the power of our structures usually bigger than the power of the noise at the scale of interest; otherwise it is not a noise but some other structures which should be analyzed as well. Finally, ICWT can be performed based on those components related to the important structures which we want to study. In Figure 4(e), the noisy structures which do not have a continuation towards the region of coarse scales are removed and scale band (4.9-500) is chosen for reconstruction. Figure 4(f) presents the result of this procedure where the original analytical signal with no noise added and reconstructed one are illustrated, respectively. All maxima are located at proper positions in the reconstruction. We need to add, actually, the average value to the ICWT in order to obtain a reconstructed signal. Also, a small part of the signal energy was removed at fine scales, the reason behind why amplitudes of the reconstructed peaks are slightly different from the true ones. This is not a deficiency of the method; rather, it is due to the choice of the scale band for reconstruction.

3. Experimental Setup

The sensor used in the experiments for this paper is a long-wave Phoenix IR camera from FLIR [13] with a thermal sensitivity of 0.025 °C. The image sequences captured by the camera have a 14-bit extended dynamic range in a 320x256 format. The video frames are acquired at a rate of approximately 30 frames per second (fps). Usually, an external black body is used to calibrate the camera to change the output in terms of temperature. In our case, the emitted infrared radiation measured by the sensor is used in the subsequent calculations (instead of temperature values). The subjects are measured from a distance of one meter due to the limitations of the existing optics of the camera. More elaborate camera lenses would dramatically increase the distance of measurements. As a baseline measurement for the heart rate measurement, a portable heart rate monitor from Polar USA, consisting of a chest strap and a watch monitor, is employed. To get a baseline measurement of breathing, the subject is requested to count the number of inhalations during the course of the experiment, pending the arrival of a piezoelectric breathing measuring device. Figure 5 illustrates the experimental setup.

4. Results and Discussion

We tested our approach on 4 healthy subjects with ages in the range of 24-35 years old, composed of three males and one female. The experiment was performed in dimly lighted room, with all subjects seated on a chair one meter away from the camera. The acquisition time for the cardiac pulse measurement was 20 sec with a frame rate of 30 fps and about 1 minute for breathing rate with frame rate of 7.5 fps. For all 4 subjects we obtained 100% accuracy for both breathing and heart rate measurements using the baseline measurements mentioned in the experimental setup as a comparison. The following figures below illustrate step-by-step the procedure. Experiments on the image sequences have shown that the Scale-2 image decomposition (Figure 4(d) provides the strongest information regarding the heart and breathing rates.
Figure 6(a) shows the time plot of the nasal area at Scale-1 decomposition. The graph is flipped (Figure 6(b)) because it is easier to take note of the peaks in the reconstructed signal later on. The sharp peaks in Figure 6(b) correspond to the inhalation of air in the breathing cycle since it is at a lower temperature and the base refers to the exhalation, where hot air is expelled. Continuous Wavelet Analysis (CWA) is applied to the 1D signal to extract the relevant information (Figure 6(c)). The peaks at the finer or small scales correspond to the noise introduced to the experiment from various sources, which can be easily removed (they do not have continuation toward bigger scale region), resulting to the wavelet plane in Figure 6(d). The breathing signal is strong enough that it can be recovered by choosing from any scale-band between 1.0-10.0. The reconstructed breathing signal, using 2.0-9.0 scale-band is shown in Figure 6(e). Breathing rate can be solved directly from the de-noised CWT plane or reconstructed waveform. There 14 peaks in 459 frames acquired with rate 7.5 fps what corresponds to 61.2 seconds. This gives us 13.7 breathing cycles per minute, which is approximately the same to the baseline measurement of 14 cycles per minute.

The computation for the heart rate is more challenging than the breathing rate since the arteries that carry the signal are well-hidden inside the body. In this paper, we focus on the carotid vessel complex of the neck but the superficial vessel complex in the forehead would work, as well. ROI-1 was chosen manually as the area which yields the strongest pulse signal and that it is common on all subjects. Figure 7(a) shows the time plot of the carotid area at Scale-2 (this scale gave us best performance). Important to note that it is impossible to obtain this waveform, similar to ECG sensors and other active modalities, using the original face thermogram [4] due to inherent noise in the system. In our approach the noise can be removed by obtaining the waveforms at different scales using multi-scale MH filters. We first decompose the thermal images at different scales and select the scale which carry the information most relevant to breathing and cardiac pulse signals than other scales. Only after these steps we construct time varying signals from points in ROI-1 and ROI-2 at this new representation followed by CWA.

Just as in the breathing measurement, CWA is applied to the 1D signal, obtained from Scale-2 (gives best performance) of ROI-1, Figure 7(b). After removing the noise at the lower scales and choosing a bandwidth in the wavelet plane (Figure 7(c)), the reconstructed signal can be obtained using inverse CWT. As before, we can compute heart beat rate directly by counting the maxima either in de-noised CWT plane or in reconstructed signal (21 peaks during 600 frames in our case). Knowing the frame rate (30 fps) of the camera, it is straightforward to compute the heart rate: 63 beats per minute (bpm). The portable heart monitor provides an average baseline measurement of 63 bpm for 20 seconds.

Conclusion

We present a novel non-contact, multi-scale wavelet-based framework for measuring vital signs such as heart rate and breathing rate. Although there are existing methods of vital signs measurement, there are several disadvantages in them that this paper is trying to solve, such as the issue of active sensors and the use of time-varying signals of original thermal images, which contain
noise and different patterns that are not relevant to vital signs measurement. Our results show that we can even extract a periodic waveform similar to contact measuring devices of heart rate such as ECG and acquire the frequency content of the signal. The future work is to develop an automatic detection of Regions of Interest (ROI) and robust tracking of features to avoid noise from motion artefacts that are detrimental to overall signal quality.

The potential application of this work is in the fields of non-contact health state monitoring and intent detection. The various scales used in extracting significant features in the face thermogram for vital signs detection can be applied to get discriminating characteristics that are useful in the two areas previously mentioned.

Acknowledgement

The authors would like to thank Dr. Edward Essock and Mr. Andrew Haun, both of the Department of Psychology and Brain Sciences, University of Louisville, for providing the IR Camera and invaluable technical assistance.

6. References

http://www.who.int/mediacentre/factsheets/fs226/en