Vital Sign Monitoring Utilizing Eulerian Video Magnification and Thermography

Bauyrzhan Aubakir, *Student Member, IEEE*, Birzhan Nurimbetov, Iliyas Tursynbek and Huseyin Atakan Varol, *Member, IEEE*

***Abstract*— In this paper we present a proof of concept for non-contact extraction of vital signs using RGB and thermal images obtained from a smart phone. Using our method, heart rate, respiratory rate and forehead temperature can be measured concurrently. Face detection and tracking is leveraged in order to allow natural motion of patients. Heart rate is estimated via processing of visible band RGB video using Eulerian Video Magnification technique. Respiratory rate and the temperature is measured using thermal video. Experiments conducted with 11 healthy subjects indicate that heart rate and respiration rate can be measured with 92 and 94 percent accuracy, respectively.**

I. Introduction

Registered nurse shortage due to rapidly ageing societies is a significant risk for the healthcare systems of developed countries [1]. Below-target nurse staffing decreases the quality of healthcare services while increasing the likelihood of adverse outcomes and patient mortality [2]. A recent study provides a closer look to the time allocation patterns of medical and surgical nurses in their routine activities [3]. The nurses in this study spent 19.3 and 7.2 percent of their nursing practice time for patient care activities and obtaining vital signs, respectively. An automated, easy-to-use and rapid system to measure vital signs such as heart rate, body temperature and respiration rate would allow more time for patient care activities.

There are several works on contactless vital signal monitoring in the literature. Utilizing a frequency domain based approach [4], researchers localized vascular maps from thermal video and estimated heart rate with 45 to 85 percent accuracy. Kwon et al. showed that it is possible to measure the heart rate of a motionless subject using a smartphone camera [5]. With regard to the respiratory rate estimation, wavelet based method and Eulerian Video Magnification (EVM) [6] were used on thermal videos in [7] and [8], respectively. EVM technique was also employed for respiration detection and rate measurement for neonatal subjects in quiet sleep condition [9]. Clinical results of non-contact respiratory rate measurement using a 2.4 GHz Doppler radar was presented in [10]. Motion noise was reduced in Doppler radar based heart rate monitoring using differential measurements with two radars [11].

Thermal camera modules that can be connected to a phone [12, 13] and smartphones with integrated thermal imaging arrays [14] are among the newest developments in mobile devices. The ability to measure vital signs via a smartphone offers the following advantages for the healthcare workers and patients. Nurses will save time by measuring multiple vital signs concurrently. Additionally, the vital signs can be uploaded from the smartphone wirelessly to an electronic medical record database saving time to the registered nurses in recording of these results as well. The patients will experience minimal discomfort thanks to the non-contact nature of the measurement.

In this paper, we present our proof of concept for vital signs measurement using thermal and RGB videos obtained from a smartphone. Hardware and software block diagram of our framework is shown in Figure 1. Specifically, spectrogram analysis of the EVM amplified RGB video from a high frame rate smartphone camera is utilized for heart rate estimation. Cycles of temperature change in the raw thermal image around the nasal region is counted for respiration rate estimation. Temperature measurement is achieved by averaging the temperature of the forehead region pixels of the thermal image. As a novelty, we leverage face detection and region of interest (ROI) segmentation to measure the vitals without restricting the natural motion of the patient and to increase the robustness of estimation.

The rest of the paper is organized as follows. Section II describes our thermographic vital sign monitoring framework using Eulerian Video Magnification. Section III presents the experimental procedures for database generation and vital measurement. Section IV shows the obtained results, while

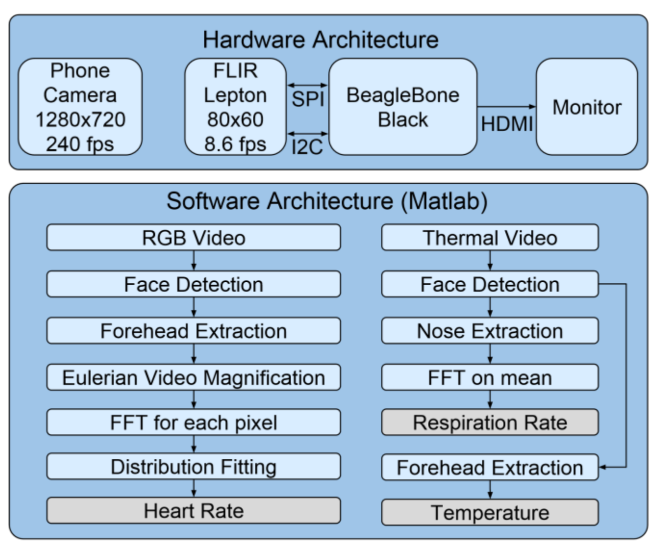

Figure 1. Hardware architecture of the RGB-Thermal data acquisition setup and the block diagram of the vital sign measurement framework.

The authors are with the Department of Robotics and Mechatronics, Nazarbayev University, 53, Kabanbay Batyr Ave, Z05H0P9 Astana, Kazakhstan. Emails: {b.aubakir, birzhan.nurimbetov, iliyas.tursynbek, ahvarol}@nu.edu.kz.
Corresponding author: Huseyin Atakan Varol.

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

Section V concludes the paper and points to the direction of future works.

II. Vital Sign Monitoring Framework

The face region in the initial frame of the RGB video is extracted using Viola-Jones algorithm [15]. In the consecutive frames the face is tracked using Kanade–Lucas–Tomasi (KLT) feature tracker [16]. From each face frame, forehead area is segmented approximately as the upper third part of the frame. EVM is applied on the extracted forehead ROI in order to amplify subtle changes in the forehead hue due to the pulse. Fast Fourier Transform (FFT) is applied on the time series signal of each pixel of the amplified ROI. The frequency of the maximum magnitude component of the spectrum for each pixel is obtained from the FFT results. A histogram of these maximum frequency values is created and non-parametric kernel density estimation is applied to compute the underlying probability density function. The frequency corresponding to the maximum value of this probability density function is outputted as the heart rate.

Similarly, the face region from the thermal video is extracted using Viola-Jones object detection algorithm. Algorithm was trained for face detection from thermal images using the database collected with the FLIR Lepton camera. Philtrum ROI from the face region is extracted. Respiration rate (RR) is obtained from the FFT of averaged values of ROI. The body temperature is computed as the mean temperature of the forehead region.

III. Implementation

A. Experimental Setup

The experimental procedures involving human subjects described in this paper were approved by the Research Committee of the School of Science and Technology of the Nazarbayev University. For monitoring heart rate, we used a phone camera with 1280x720 resolution and 240 Hz frame rate. The high frame rate was chosen in order to get results with higher precision. In order to get a reliable estimate, videos were recorded for a period of 30 seconds. The videos were then transferred to a personal computer with Windows 10 operating system and processed using MATLAB.

For monitoring respiratory rate and body temperature we used a FLIR Lepton long range infrared camera with 80x60 resolution, 8.6 Hz frame rate and 50 mK thermal sensitivity. The thermal camera is connected to a BeagleBone Black single board computer. SPI communication protocol is used to send the raw data from the thermal imaging camera, while I2C interface is used for calibration. BeagleBone is connected to a monitor via HDMI port for a convenient user interface during data collection. The raw data for each subject was again recorded as 30 second videos, which were transferred to a computer and processed using MATLAB.

Both phone with a camera and thermal camera with BeagleBone Black are enclosed in a 3D printed enclosure and mounted to a tripod for maintaining a stable pose (see Figure 2). The subjects were positioned approximately 30-40 cm distant from the data acquisition setup in full-face view (see Figure 3). A white curtain was placed behind the subject to ensure a uniform background for visible and thermal band images. RGB and thermal videos were captured simultaneously in an environment with daylight and no flickering light sources.

B. Face Detection and ROI Segmentation

We implemented face detection and tracking in order to observe the vital signs in the ROI from the obtained RGB and thermal videos (see Figure 4). For the heart rate and temperature estimation our focus lies in the forehead region. It has a dense vascular system containing supraorbital and supratrochlear arteries. It is broad and flat compared to the other regions of the face, usually uncovered and easily detectable. For the estimation of the respiration rate we used nasal region as ROI. In that region of the thermal video, changes of the temperature between inhalation and exhalation can be observed clearly.

The above mentioned parts of the face were detected using the MATLAB implementation of the Viola-Jones

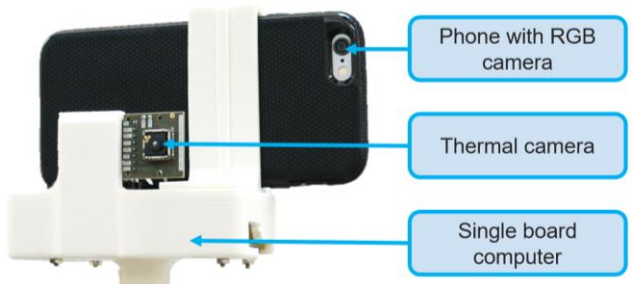

Figure 2. Hardware components of the experimental setup.

Figure 3. Relative positioning of the data acquisition setup with respect to the experimental subject.

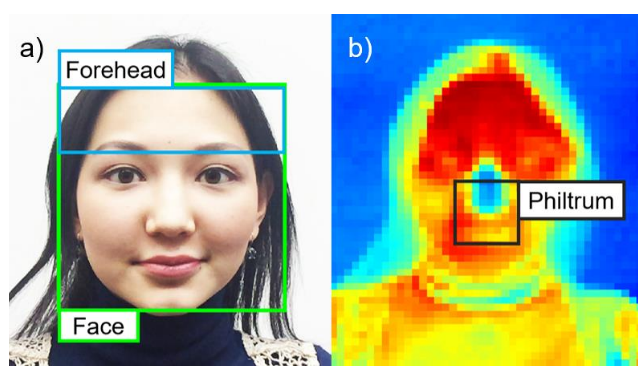

Figure 4. Results of the face detection and the ROI extraction procedure: RGB image with marked face and forehead ROIs (a) and thermal image with marked nasal ROI (b).

object detection framework. Later, the detected parts were tracked using the KLT feature tracker such that the same regions are processed in each frame eliminating effects of the subject's motion.

Face detection classifier for RGB images is available in MATLAB. It was used to detect a face in the RGB video for heart rate estimation. However, this classifier does not work properly with the thermal images since it was trained with visible band RGB data. For this purpose, we created a database of 47 thermal videos of different persons for training our own classifier. 12 females and 35 male subjects with ages ranging from 18 to 56 participated in the thermal face video database generation. In order to train the face detection algorithm, positive (face present) and negative (face not present) thermal images were required. From the thermal face video database, we took three frames of each person randomly, resulting in 141 positive images for our classifier. Negative images were taken from 40 different videos of various body parts and other objects with a range of temperatures and different poses. Ten frames were taken from each video adding 400 negative images to the database for classifier training. Faces were manually selected on the positive images using MATLAB's Training Image Labeler application. This data with specific regions of interest was used for training our classifier for thermal face detection. Two thirds of the data were used for classifier training and one third was set aside for testing. Trained classifier achieved 97.8 percent accuracy. All the faces were detected correctly with only few instances of false positives. Being able to detect faces on thermal video allowed us to extract the nasal and forehead regions for respiration rate and temperature estimation, respectively.

C. EVM Based Heart Rate Monitoring Using RGB Video

After performing face detection and ROI segmentation in the RGB video stream, we perform EVM procedure on the forehead of the subject. Assuming that the heart rate is between 0.6 and 3 Hz corresponding to 36 and 180 beats per minute (bpm), we amplified the video 20 times for this frequency range. After this procedure heart's diastole (relaxation) and systole (contraction) can be observed (see Figure 5a and 5b). The amplified video was converted to hue-saturation-value color model. FFT was performed on the time series signals for each pixel in the hue channel (corresponding to 7200 samples for a 30 second video). The frequency component with the highest magnitude was recorded for each spectrogram and combined into a histogram, where the highest bar indicated the raw heart rate estimate. To obtain results with more precision, we fit probability distribution using kernel density estimation method and select its maximum as the heart rate. The histogram and the fitted kernel distribution are shown in Figure 5c.

D. Respiratory Rate Monitoring Using Thermal Video

After performing face detection on the thermal video, software extracts respiratory rate. In this case, it is not necessary to perform Eulerian Video Magnification, since the changes can be observed in original video. From the detected face we extract nasal area and measure the average temperature of that area. Depending on the nasal area anatomy and temperature of the environment, philtrum temperature change could be in the range of 0.25 K to 2 K. After measuring the means, we obtain FFT of the data and show the strongest frequency as respiration rate. Thermal

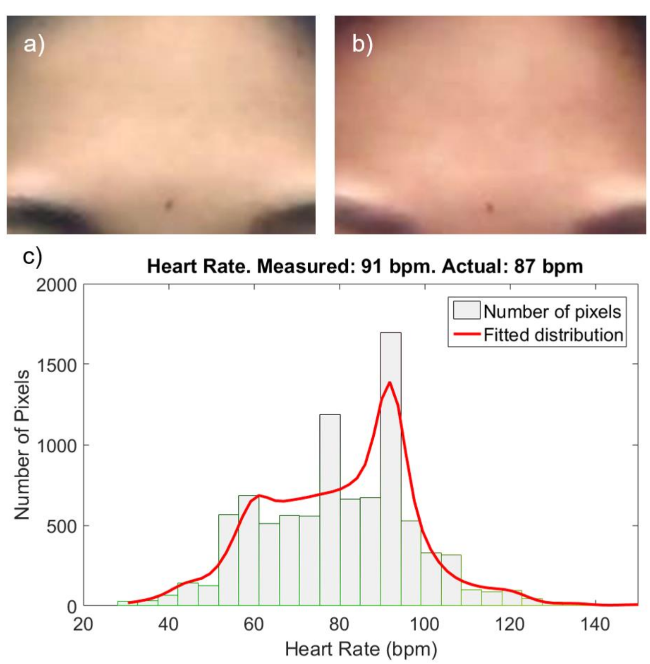

Figure 5. EVM amplified RGB images during the diastole (a) and systole (b) parts of the heart cycle. Histogram of the maximum frequency components of each forehead region pixel from the RGB video (c).

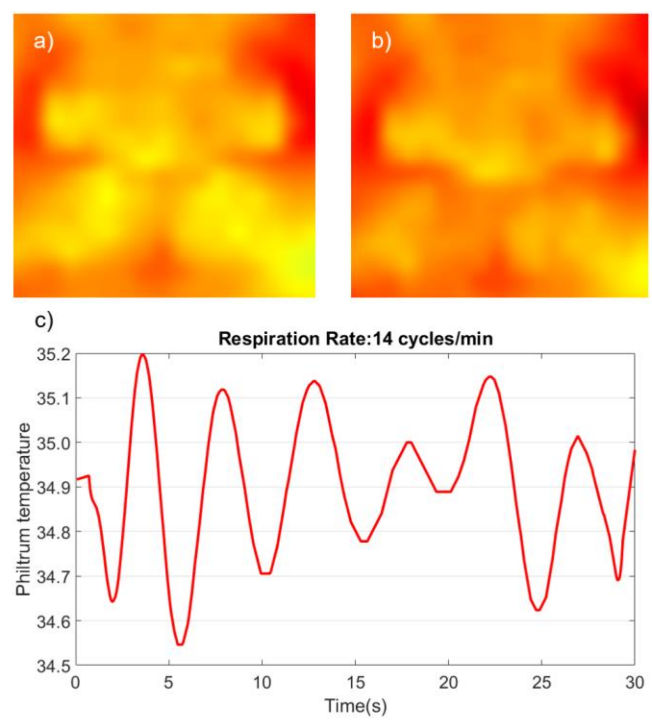

Figure 6. Thermal image of the nasal area obtained during inhalation (a), and exhalation (b). Positive peaks of the philtrum temperature change are used for respiratory rate estimation (c).

images of the nasal region during inhalation and exhalation are shown in Figure 6a-b. The average temperature of the nasal region for a 30 second trial is shown in Figure 6c.

IV. Results

Using the vital sign monitoring framework, we acquired the heart rate of 11 subjects and calculated the accuracy by comparing it to the values obtained from Sigma Sport PC 10.11 heart rate monitor. The accuracy was defined as

$$Accuracy = \left(100 - 100\frac{|HR_{monitor} - HR_{video}|}{HR_{monitor}}\right)$$

where $HR_{monitor}$ and HR_{video} are the heart rates measured using heart rate monitor and our RGB video amplification method, respectively. The average accuracy for heart rate estimation was around 92 percent. The results of heart rate measurement experiments are summarized in Table I.

The respiratory rate was also measured for these 11 subjects. Estimated values were compared to the actual values of respiratory rate, which were measured manually. The average accuracy for respiration rate was 94 percent. The results of respiratory rate measurement experiments and the average forehead temperatures are given in Table II.

V. Conclusion

In this paper we demonstrated a proof of concept for an "in-phone" contactless vital sign monitoring system. Our method takes an RGB video from a phone camera and a thermal video from an infrared camera and obtains heart rate, respiratory rate and body temperature using advanced signal processing techniques. The results of our experiments show that our method can be used for vital sign monitoring with over 90 percent accuracy. This method could be used by medical staff in order to measure vital signs in an efficient and automated manner. In future we plan to implement our framework in smartphones. We also speculate that the vital sign measurement system can be implemented only using thermal imaging with the introduction of higher resolution and frame rate cameras.

TABLE I. Average Pulse Rate Measured by the Heart Rate Monitor and Extracted From the EVM Amplified RGB Video

Subject	Actual Heart Rate	Estimated Heart Rate (Accuracy %)
1	64	69.0 (92.1)
2	87	91.8 (94.5)
3	92	86.0 (93.5)
4	73	90.8 (75.6)
5	89	79.6 (89.4)
6	66	68.0 (97.0)
7	106	112.0 (94.3)
8	80	81.9 (97.6)
9	65	76.0 (83.1)
10	66	64.0 (97.0)
11	116	112.4 (96.9)

TABLE II. Vital Signs Obtained From Thermal Video: Actual and Estimated Respiration Rate (RR) and Forehead Temperature

Subject	Actual RR	Estimated RR (Accuracy %)	Forehead Temperature (ºC)
1	13	14 (92)	36.15
2	12	12 (100)	35.95
3	15	15 (100)	36.85
4	18	17 (94)	34.95
5	17	17 (100)	35.10
6	15	14 (93)	36.55
7	26	26 (100)	36.10
8	26	25 (96)	36.10
9	18	18 (100)	34.90
10	14	19 (64)	36.70
11	27	28 (96)	37.00

References

[1] S. P. Juraschek, X. Zhang, V. Ranganathan, *et al.*, "United States registered nurse workforce report card and shortage forecast," *American Journal of Medical Quality,* vol. 27, pp. 241-249, May 2012.

[2] J. Needleman, P. Buerhaus, V. S. Pankratz, *et al.*, "Nurse staffing and inpatient hospital mortality," *The New England Journal of Medicine,* vol. 364, pp. 1037-1045, Mar 2011.

[3] A. Hendrich, M. P. Chow, B. A. Skierczynski, *et al.*, "A 36-hospital time and motion study: How do medical-surgical nurses spend their time?," *The Permanente Journal,* vol. 12, pp. 25-34, Summer 2008.

[4] T. R. Gault and A. F. Farag, "A fully automatic method to extract the heart rate from thermal video," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Portland, OR, 2013, pp. 336-341.

[5] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, San Diego, CA, 2012, pp. 2174-2177.

[6] H. Y. Wu, M. Rubinstein, E. Shih, *et al.*, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics,* vol. 31, Jul 2012.

[7] C. S. Yu, H. Rara, and A. A. Farag, "Non-contact, wavelet-based measurement of vital signs using thermal imaging" in *First International Conference on Graphics, Vision, and Image Processing (GVIP)*, Cairo, Egypt, 2005.

[8] S. L. Bennett, R. Goubran, and F. Knoefel, "The detection of breathing behavior using Eulerian-enhanced thermal video," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* vol. 2015, pp. 7474-7, Aug 2015.

[9] N. Koolen, O. Decroupet, A. Dereymaeker, *et al.*, "Automated respiration detection from neonatal video data," in *International Conference on Pattern Recognition Applications and Methods*, 2015, pp. 164-169.

[10] A. D. Droitcour, O. Boric-Lubecke, and G. T. A. Kovacs, "Signal-to-noise ratio in Doppler radar system for heart and respiratory rate measurements," *IEEE Transactions on Microwave Theory and Techniques,* vol. 57, pp. 2498-2507, Oct 2009.

[11] R. Fletcher and J. Han, "Low-cost differential front-end for Doppler radar vital sign monitoring," in *IEEE International Microwave Symposium*, Boston, MA, 2009, pp. 1325-1328.

[12] FLIR. (2016, March 2016). *FLIR ONE*. Available: http://www.flir.com/flirone/

[13] S. Thermal. (2016, March 2016). *CompactXR: Extra-range thermal imaging camera designed for your smartphone*. Available: http://www.thermal.com/products/compactxr

[14] Caterpillar. (2016, March 2016). *CAT S60 Smartphone*. Available: http://www.catphones.com/en-gb/phones/s60-smartphone

[15] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision,* vol. 57, pp. 137-154, May 2004.

[16] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, 1994, pp. 593-600