Sensitivity of the Contactless Videoplethysmography-Based Heart Rate Detection to Different Measurement Conditions

Ennio Gambi, Manola Ricciuti, Susanna Spinsante
Dipartimento di Ingegneria dell'Informazione
Università Politecnica delle Marche
Ancona, Italy 60131
Email: {e.gambi, m.ricciuti, s.spinsante}@staff.univpm.it

***Abstract*—Technologies for contactless Heart Rate measurement support the progress in the diagnostic and healthcare fields, opening new possibilities even for everyday use at home. Among them, Videoplethysmography based on the Eulerian Video Magnification method has been already validated as an effective alternative to traditional, but often bulky, Electrocardiographic acquisitions. In this paper we study the influence of different measurement parameters on the Heart Rate estimation, in order to assess the reliability of the Videoplethysmography detection method under varying conditions, like different dimensions and positions of the processed regions of interest, pyramidal decomposition levels, and light conditions.**

I. Introduction

The possibility to unobtrusively measure biomedical parameters may be of great importance in healthcare, as it can enable diagnosis and follow-up of different medical issues in very severe conditions, like those of burned patients, or when newborns in incubators are to be monitored. Among the parameters that can be contactlessly detected, the Heart Rate (HR) is of interest, as the number of times the heart beats per minute may be affected by disease, the level of activity or stress, and by the subject's age. The HR of a subject may change along time depending on the natural aging process, but variations may indicate also possible risky conditions making the heart more susceptible to failure [1].

The gold standard method for heartbeat detection adopts an Electrocardiograph (ECG), an expensive and non comfortable medical device, the use of which is typically limited to clinical premises [2], [3]. Commercial certified ECGs are available too, that provide both HR estimation and blood pressure values. Nowadays, wearable devices like smart watches or wrist bands claim to provide a number of parameters, among which the HR value too, but reliability and accuracy of HR estimation may vary across different brands and platforms [4]. In all the above mentioned cases, the technology used for HR detection requires a direct contact of the device and the human body, in order to extract the voltage or the Photoplethysmographic (PPG) signal [5] to process.

PPG-based methods for HR detection exploit light signals and two possible modalities, i.e. reflection or transmission. The PPG signal is obtained from the spectra of light reflected from (or transmitted through) body tissues, once the skin is illuminated by a LED and the amount of light reflected (or transmitted through) is measured. Medical devices based on PPG technology, such as pulse oximeters [6], are widely used in different healthcare applications. PPG has been also proposed for the analysis of Heart Rate Variability (HRV) [7], [8] as an alternative to ECG, even if the method may be less accurate, due to the difficult detection of peaks in the measured signal. Solutions to measure the HR without body contact offer technological and scientific progress in the diagnostic and healthcare fields, opening new possibilities even for everyday use at home, thanks to a simplified HR acquisition procedure. As a consequence, we are interested on the evaluation of the HR detection performed by video processing, which requires collecting a Videoplethysmographic (VPG) signal [9] of the subject to monitor, through a contactless procedure. As a matter of fact, the blood flowing in the tissues varies according to the subject's HR, and causes, among others, extremely small variations of the skin color, that are not normally perceptible to the human eye. However, through proper VPG signal processing techniques, based on the Eulerian Video Magnification (EVM) algorithm [10], it is possible to process the Luminance (Y) component of the captured VPG and obtain, as an indirect measurement, the HR value. Several processing steps have to be applied on the Y component of the captured video frames, in order to extract the VPG signal. At each step, different parameters are set, and they may influence the final HR detection.

Following a previous work of the same authors [11], in which the VPG-based contactless HR estimation was validated against the gold-standard measurement procedure using a Holter device, in this paper we study the influence of different measurement parameter configurations on the HR estimation, in order to assess the reliability of the detection method. Using the same dataset of video sequences acquired through the RGB camera of a Kinect device [11], several configurations of the measurement parameters are considered, and the effects of light variations are analyzed too.

The paper is organized as follows: in Section II a summary of the processing steps leading from the acquired video

ISBN 978-90-827970-1-5

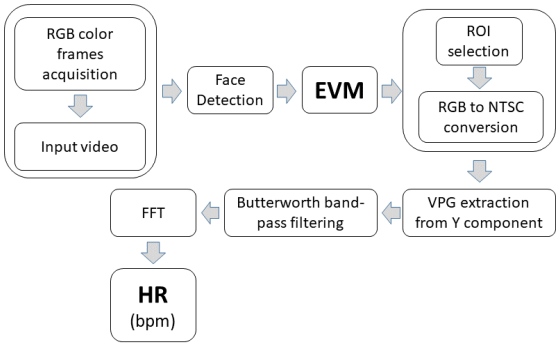

Fig. 1. Main scheme of the proposed system for the computation of the subject's HR from RGB video.

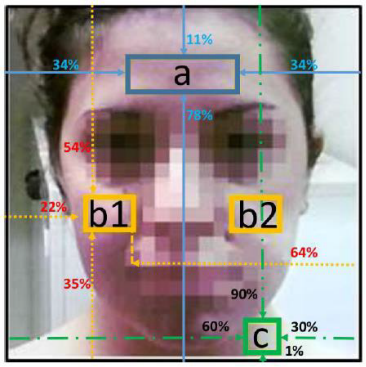

Fig. 2. ROIs identification according to [11] (face details removed for privacy reasons): **a** is ROI_f, $\mathbf{b}_1$ and $\mathbf{b}_2$ are ROI_c, and **c** is ROI_n.

RGB frames to the VPG signal is provided, focusing on the parameters that will be involved in the assessment of the HR detection method. In Section III, the effects of different parameter values selection on the HR detection are presented through experimental evaluation, and a thorough discussion of the results is provided in Section IV. Finally, Section V draws the main conclusion of this work.

II. HR EXTRACTION FROM VIDEO PROCESSING

Video processing for HR extraction takes as input RGB frames of a human face. The area related to the human face is detected and processed by the EVM algorithm [10], which enhances the small changes in RGB levels due to blood flow. Then, limited areas of the face and neck are selected through a Region Of Interest (ROI) extraction step. Following an RGB-to-YIQ color space conversion, the main frequency component of the Luminance (Y) signal within the ROI is computed through the Fast Fourier Transform (FFT), and the HR value is the output of the system. A schematic view of the method is given in Fig. 1, where the different steps can be identified.

A. VPG processing steps and parameters

After the RGB video capture, that takes place at a maximum frame rate of 30 fps for a total time of 40 s at each acquisition, the well-know Viola-Jones algorithm [12] is applied for face detection. This way, the relevant area of the frame, i.e. the one that includes face and neck, is selected and tracked along subsequent frames by the Kanade-Lucas-Tomasi (KLT) algorithm [13], [14]. Following this procedure, only the relevant parts of the original RGB frames are selected, thus reducing the computational burden of the following EVM operations.

The EVM method used to magnify the skin color changes due to the heart beats requires the proper configuration of some parameters, namely:

- the color amplification level, denoted as α;
- the Gaussian piramidal decomposition level, denoted as L, and
- the frequency bandwidth of the ideal band-pass filter, defined by f_L and f_H.

According to the EVM method, in order to extract the HR value from the VPG signal, each RGB frame is converted to the YIQ color space, and the Y component is selected for further processing, consisting in: a filtering operation by a third order Butterworth filter featuring the same cut-off frequencies of the ideal band-pass filter used in the EVM, and an FFT stage on the filtered Y component. Finally, the subject's HR in bpm (beats per minute) is obtained by extracting the peak frequency component and multiplying it by a factor of 60. In our previous work [11], the above mentioned parameters were set heuristically to the following values:

- $\alpha = 100$, to have a good compromise between signal amplification and the presence of artifacts;
- $L = 6$;
- $f_L = 50/60$ Hz and $f_H = 174/60$ Hz.

Additionally, based on empirical outcomes, the minimum size of the frame area to perform the face detection process was set to 280×280 pixels, and the forehead, cheeks, and neck ROIs (denoted as ROI_f, ROI_c, and ROI_n) were chosen equal to 90×30, 40×30, and 28×26 pixels, respectively. A pictorial representation of the ROIs is shown in Fig. 2.

B. Experimental data acquisition

The RGB frames processed by the EVM method are acquired through a Microsoft Kinect v2 sensor: they are stored as bitmap images, with a resolution of 1920×1080 pixels at 30 fps frame rate, using the software tool developed by the authors [15]. A number of 20 young subjects (10 males and 10 females) participated to the validation tests (age: (22.50 $\pm$ 1.57) yrs; height: (173 $\pm$ 10) cm; weight: (62.80 $\pm$ 9.52) kg), and for each subject five video sequences were acquired, generating a total number of 100 sequences. For practical reasons, subjects were facing the Kinect while sitting on a chair at a distance of about one meter, as shown in Fig. 3. This accounts for the fact that the horizontal field of view of the device is $(70 \pm 5)^\circ$ and the vertical field of view is $(60 \pm 5)^\circ$. The acquisition time length has been set to 40 s.

III. VARIATION OF MEASUREMENT PARAMETERS

As extensively discussed in [11], the performance of the proposed contactless and VPG-based HR detection method are comparable to those obtainable from the gold standard

ISBN 978-90-827970-1-5 © EURASIP 2018

Fig. 3. The experimental data acquisition setup [11].

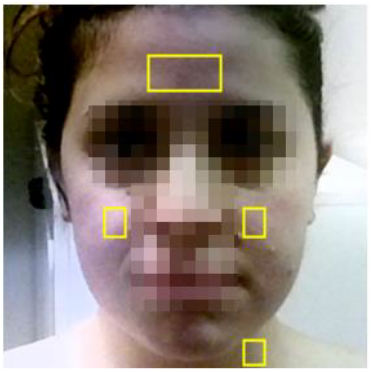

Fig. 4. Reduced ROI dimensions.

Holter device. However, as there are many parameters involved in the different processing steps, it is important to check the robustness of the HR estimation with respect to possible (controlled or uncontrolled) variations of those parameters. To this aim, in this Section we investigate how the estimated HR values may vary, due to different dimensions of the selected ROIs, possible displacement of the selected ROI with respect to its ideal position, and variations of the environmental light conditions.

A. Variation of the ROI dimensions

In a first series of tests, the dimensions (in pixels) of the different ROIs are reduced, and the error in the estimated HR with respect to the HR computed over the original ROIs in [11], is evaluated. Let us define the total original ROI (ROI_T) as the sum of the areas of all the original ROIs involved: $ROI_T = ROI_f + 2 \cdot ROI_c + ROI_n$. ROIs with reduced dimensions are identified by a superscript, according to Table I. In Experiment 1, the ROI_T^1 amounts to 38.7% of ROI_T, whereas in Experiment 2, the ROI_T^2 amounts to 44.5% of ROI_T. It is important to outline that ROI dimensions have to be chosen according to the area of the frame selected for the face detector, which amounts to 280×280 pixels, out of 1920×1080 pixels in a frame, in our case. So the ROI dimensions are set proportionally to the face detector area. On average, the distance of each ROI side from the parallel face detector boundary has been increased by 10%, resulting in the reduction of the ROI areas specified in Table I, and illustrated by a sample in Fig. 4, compared to Fig. 2.

In order to quantify the HR estimation error due to the use of reduced ROI dimensions, the following definition of *relative error* is given:

$$\epsilon_r = \frac{\epsilon_a}{HR_{avg}} \cdot 100 \qquad (1)$$

where HR_{avg} denotes the average value of the collected measurements, and ϵ_a is the absolute error.

Referring to the settings described in Table I, we obtained $\epsilon_r^1 = 6.55\%$ in Experiment 1, and $\epsilon_r^2 = 6.52\%$ in Experiment 2, respectively. Keeping the same configuration of the other parameters used in [11] ($\alpha = 100$, $L = 6$, $f_L = 50/60$ Hz, and $f_H = 174/60$ Hz), the relative error increases from 3.40% originally obtained through the *unsupervised* approach (according to which no a priori knowledge about the subject's health status is available), to around 6.5% in both Experiment 1 and 2.

B. Variation of the Gaussian pyramidal decomposition level

In order to reduce the relative error resulting from the reduction of the ROI dimensions, in a second series of tests we consistently reduced the L parameter from the original value of 6 down to 4. In fact, the Gaussian pyramidal decomposition implies a sub-sampling of the frame and a decrease in its resolution; by limiting the decomposition level to $L = 4$ we improve the trade-off between ROI dimension and decomposition (i.e. resolution). In this case, by processing again all the 100 VPG sequences collected, for the ROI dimensions specified in section III-A, we obtained a minimum relative error $\epsilon_r^L = 4.74\%$, thus denoting an improvement with respect to Experiment 1 and 2.

C. Variation of the neck ROI position

An additional series of tests aimed at evaluating the effects of ROIs displacement on the HR estimation. In Table 2 of [11], it is shown how the use of the single ROI_n in the *supervised* mode (filter parameters are adjusted according to the known condition of the subject, sporty or sedentary) provides a good HR estimation, with $\epsilon_r^S = 2.70\%$. In the unsupervised mode, more interesting in the perspective of a practical implementation of the method, $\epsilon_r^U = 6.69\%$. This result is quite reasonable, as the neck area includes the carotid arteries, over which the EVM provides good outcomes. However, placing the ROI_n on the left or right side of the neck, may affect the result. In fact, the left carotid artery (right side of the neck area in the video) originates directly from the aorta, thus it is reasonable to expect a stronger VPG outcome, whereas the right carotid artery (left side of the neck area in the video) is the brachiocephalic one [16]. In order to test this hypothesis and check the effect of the ROI_n displacement on the HR estimation, the dimension of the ROI_n and the setting of the EVM parameters were restored to the original values

ISBN 978-90-827970-1-5

TABLE I
REDUCED DIMENSIONS OF THE ROIS USED IN EXPERIMENTS. "ORIGINAL" DENOTES THE SETTINGS USED IN [11]

Test	forehead area	cheek area	neck area	total area
Original	$ROI_f = 90 \times 30 = 2700$ px	$ROI_c = 40 \times 30 = 1200$ px	$ROI_n = 28 \times 26 = 728$ px	$ROI_T = 5828$ px
Exp. 1	$ROI_f^1 = 56 \times 22 = 1232$ px	$ROI_c^1 = 22 \times 16 = 352$ px	$ROI_n^1 = 16 \times 20 = 320$ px	$ROI_T^1 = 2256$ px
$\Delta_1\%$	45.6%	29.3%	43.9%	38.7%
Exp. 2	$ROI_f^2 = 70 \times 22 = 1540$ px	$ROI_c^2 = 22 \times 22 = 484$ px	$ROI_n^2 = 22 \times 26 = 572$ px	$ROI_T^2 = 2596$ px
$\Delta_2\%$	57%	40.3%	78.5%	44.5%

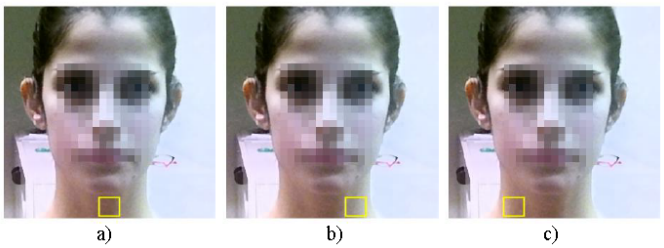

Fig. 5. Neck ROI displacement in 3 steps: a) C, b) R - corresponding to anatomic left side, and c) L - corresponding to anatomic right side.

TABLE II
RELATIVE HR ESTIMATION ERROR (AVERAGE OVER 100 SEQUENCES) ACCORDING TO ROI_n LOCATION

ϵ_r^R	ϵ_r^C	ϵ_r^L	ϵ_r^U
6.40%	14.01%	18.09%	6.69%

of [11] (see also Table I in this paper), but three different positions were considered: right (R - that corresponds to the anatomic left side of the neck), center (C), and left (L - anatomic right side), as shown in Fig. 5. By repeating the experimental tests with the ROI_n located in the three different positions, the results provided in Table II were obtained. A slight improvement of the relative HR estimation error was obtained on the right side location, i.e. the anatomic left side, as expected, with $\epsilon_r^R = 6.40\%$.

Additional tests were executed by reducing the ROI_n dimension to 22×26 pixels (ROI_n^2), and the pyramid decomposition level accordingly ($L = 4$). Then, eigth different positions were tested, moving the ROI_n^2 from left, to center, to right in the video frame, as shown in Fig. 6. This way, in the two rightmost positions of the video, the relative error was decreased, down to 5.51%, as detailed in Table III (ϵ_r^8); however, the remaining positions did not provide any improvement but, on the contrary, a relevant increase of the relative HR estimation error.

D. Variation of the light conditions

As a final experiment, the effect of light variations in the acquired video sequences was tested. To such an aim, an artificial increase of the brightness factor (BF) was applied on half of the frame area according to the histogram slide method [17], with two different values: $BF = 25$ and $BF = 50$. In

TABLE III
RELATIVE HR ESTIMATION ERROR (AVERAGED OVER 100 SEQUENCES) ON EIGHT ROI_n^2 LOCATIONS

ϵ_r^1	ϵ_r^2	ϵ_r^3	ϵ_r^4
17.65%	13.91%	14.30%	15.51%
ϵ_r^5	ϵ_r^6	ϵ_r^7	ϵ_r^8
14.80%	13.59%	5.93%	5.51%

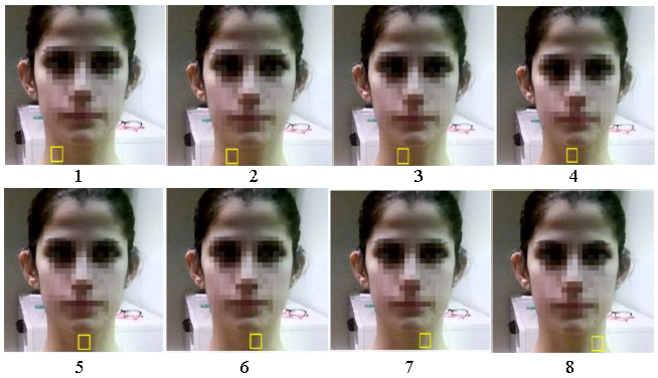

Fig. 6. Neck ROI displacement in 8 steps, from left side (anatomic right one) to right side (anatomic left one).

Fig. 7 b) and c), an example of artificially illuminated video frame is shown for $BF = 25$ and $BF = 50$, respectively, whereas Fig. 7 a) shows the original frame.

The impact of light variations was estimated by considering either three different positions of the ROI_n (L, C, and R) as in section III-C, and the ROI_T. The results obtained are summarized in Table IV, for $BF = 25$ and $BF = 50$.

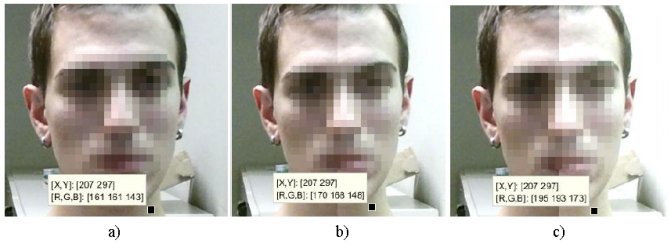

Fig. 7. Light variations in the video frame: a) original frame, b) right half frame artificially illuminated with $BF = 25$, c) right half frame artificially illuminated with $BF = 50$.

ISBN 978-90-827970-1-5

TABLE IV
RELATIVE HR ESTIMATION ERROR FOR $BF = 25$ AND $BF = 50$, EVALUATED ON ROI_n^R, ROI_n^C, ROI_n^L, AND ROI_T

BF	ϵ_r^R	ϵ_r^C	ϵ_r^L	ϵ_r^{TOT}
25	11.95%	12.21%	17.38%	6.79%
50	12.87%	13.05%	17.31%	6.19%

IV. DISCUSSION OF RESULTS

In section III-A we found that the variation of the ROI dimensions does not affect substantially the HR estimation performance. In fact, for different ROI_T^1 and ROI_T^2 dimensions, we obtained $\epsilon_r^1 \approx \epsilon_r^2$. The possibility of reducing the dimension of the ROI without affecting the HR detection performance may be important, as it can allow to discard parts of the face area in the frame that result in a noisy detection, like those corresponding to beard, moustache, glasses, or hair covering the face skin, or to apply the algorithm even on subject with a thinner neck. Considering the total amount of 100 signals processed, and the fact that the relative percentage error is acceptable when below 10%, the resulting 4.74% relative error associated to the reduction of the ROI dimensions shows the accuracy of the approach used.

However, the reduction of the single ROI dimensions, and consequently the decrease of the total ROI area, does not bring any relevant improvement to the HR detection process compared to [11], if not applied jointly with a reduction of the Gaussian pyramidal decomposition level, from $L = 6$ to $L = 4$. Smaller ROIs are effective if the image resolution is consistently adjusted by proper selection of L. The same is confirmed by the results presented in section III-C where only after the reduction of L, the relative HR estimation error was improved by using a smaller ROI_n area, on the rightmost positions of the neck area.

Finally, as shown in section III-D, the artificial variation of the frame luminance generally impacts the HR detection performance in a negative fashion, with an increase of the relative error, irrespective of the ROI considered. However, while the ROI_n results to be unusable in any location analysed, the relative error evaluated on ROI^T remains acceptable, being well under the 10% threshold assumed as a reference in the literature [18].

V. CONCLUSION

The results presented in this paper show that the VPG-based HR estimation approach validated in our previous study [11] may be further improved by the proper translation of the neck ROI. In fact, setting the accurate positioning of the ROI_n in the right side of the neck (anatomic left side), the relative HR estimation error was reduced. The performance deterioration due to alteration of brightness suggests that nonuniform light variations may be harmful and should be avoided by properly controlling the acquisition setup. Even the reduction of individual ROI areas does not lead to improvements, unless a lower level of the EVM spatial decomposition (from $L = 6$ to $L = 4$) is chosen. This implies that a smaller ROI leads to an acceptable result if the image resolution is correspondingly sub-sampled at a decomposition level not higher than 4. Future work includes the execution of a significant number of tests over more subjects, in different conditions of natural light intensity properly measured, to evaluate the robustness of the HR extraction, and comparison to other systems.

REFERENCES

[1] J. Strait and E. Lakatta, "Aging-associated cardiovascular changes and their relationship to heart failure," *Heart failure clinics*, vol. 8, no. 1, pp. 143–164, 2012.

[2] D. J. Ewing, J. M. Neilson, and P. Travis, "New method for assessing cardiac parasympathetic activity using 24 hour electrocardiograms," *British Heart Journal*, vol. 52, no. 4, pp. 396–402, 1984.

[3] T. G. Farrell, Y. Bashir, T. Cripps, M. Malik, J. Poloniecki, E. D. Bennett, D. E. Ward, and A. J. Camm, "Risk stratification for arrhythmic events in postinfarction patients based on heart rate variability, ambulatory electrocardiographic variables and the signal-averaged electrocardiogram," *Journal of the American College of Cardiology*, vol. 18, no. 3, pp. 687–697, 1991.

[4] M. Wallen, S. Gomersall, S. Keating, U. Wislff, and J. Coombes, "Accuracy of heart rate watches: Implications for weight management," *PLoS ONE*, vol. 11, no. 5, 2016.

[5] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light." *Optics express*, vol. 16, pp. 21 434–45, 2008 Dec 22 2008.

[6] S. V. More and P. C. Chaudhari, "Development of non-invasive diagnostic tool for diseases using photo plethysmography," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, March 2016, pp. 1499–1503.

[7] J. Kranjec, S. Begu, G. Gerak, and J. Drnovek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control*, vol. 13, pp. 102 – 112, 2014.

[8] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, *Comparison of Heart Rate Variability from PPG with That from ECG*. Cham: Springer International Publishing, 2014, pp. 213–215.

[9] R. R. Fletcher, D. Chamberlain, N. Paggi, and X. Deng, "Implementation of smart phone video plethysmography and dependence on lighting parameters," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 3747–3750.

[10] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 65:1–65:8, July 2012.

[11] E. Gambi, A. Agostinelli, A. Belli, L. Burattini, E. Cippitelli, S. Fioretti, P. Pierleoni, M. Ricciuti, A. Sbrollini, and S. Spinsante, "Heart rate detection using microsoft kinect: Validation and comparison to wearable devices," *Sensors*, vol. 17, no. 8, p. 1776, 2017.

[12] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[13] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679.

[14] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University, Technical Report CMU-CS-91-132, April 1991.

[15] Kinect complete viewer v2.0. [Online]. Available: http://www.tlc.dii.univpm.it/blog/databases4kinect

[16] J. Gürbüz, S. Çavdar, and Ö. Özdoğmuş, "Trifurcation of the left common carotid artery: A case report," *Clinical Anatomy*, vol. 14, no. 1, pp. 58–61, 2001.

[17] S. E. Umbaugh, *Digital image processing and analysis: human and computer vision applications with CVIPtools*. CRC press, 2016.

[18] I. Bosi, C. Cogerino, and M. Bazzani, "Real-time monitoring of heart rate by processing of microsoft kinect 2.0 generated streams," in *2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, July 2016, pp. 1–6.

ISBN 978-90-827970-1-5 © EURASIP 2018