Noncontact Automatic Heart Rate Analysis in Visible Spectrum by Specific Face Regions

Dragos Datcu, Marina Cidota, Stephan Lukosch, Leon Rothkrantz

Abstract: *The current paper presents a comparative study on the influence of different face regions for contactless extraction of the heart rate by computer vision in visible spectrum. A second novelty of our research is the use of Active Appearance Models for computing the shape of the face and of the facial features. Following an experimental setup, we determine that forehead and cheek face regions are more relevant for computing the heart rate. This outcome leads to an optimized face scanning method, faster processing times and better pulse detection results. The findings were implemented in an automatic system prototype for noncontact face analysis. Our methods were tested and validated using video recordings of people in laboratory setup.*

Key words: *Heart rate detection, Face Analysis, Photoplethysmography.*

INTRODUCTION

Non-contact analysis of physiological indicators has a major impact in several application domains making use of live monitoring of the people's faces, in realistic scenarios involving working, playing and resting.

Photoplethysmography – PPG [1] represents a noncontact, non-invasive and low-cost method that makes use of the variations of the transmitted or reflected light to determine the cardiovascular blood volume. Considering normal ambient light as the illumination source allows for using video cameras and even regular webcams as heart rate detection sensors.

In a previous study we have investigated the influence of the sampling rate on the heart rate detection and concluded that there is no actual performance gain in using video frame rates higher than 15fps [5]. By using an industrial quality level camera, we were able to reduce the size of the analysis window from 30 seconds of the previous best model in the literature [13] down to 5 seconds. The current paper presents a comparative study on the influence of different face regions on the performance of pulse detection. The motivation for this research comes from the need to optimize the face scanning process with the purpose to reduce the computation time and to increase the accuracy of the results. In the paper we will mention which face regions are the best and how different are these compared to the whole face, in the context of analysing the heart rate based on pixel RGB patterns. In addition to the analysis on separate face regions, another novelty of this paper is the integration of Active Appearance Models – AAM [7] for the accurate face and facial feature shape extraction. Knowing precisely the location of facial features, as opposed to the rough estimation of the location of the face region, helps in correctly sampling skin data and to better handle situations of head rotation.

The following section of the paper presents a literature survey on the heart rate detection. The next section describes our models for face detection, face shape extraction and for heart rate detection. Then, the automatic heart rate detection system is tested and discussed in the next section. The last section discusses the conclusions of our research and presents the future work.

RELATED WORK

Grenecker [10] was the first to investigate the detection of the physiological measurements of pulse and respiration rate in a contactless manner. The method was proposed in 1997, used radar technology and worked at a distance of maximum 10 meters. Recent studies [12][13] use information on the face area which is automatically detected in video sequences, to define a region of interest from where to segment skin

region and further on to sample red, green and blue channels data. Other works make use of semiautomatic methods involving manual annotation of the region of interest in the first frame and an update step on this region using conditional density propagation [9].

The plethysmographic signal can be separated from other sources of fluctuations in light by using Independent Component Analysis – ICA [3][6]. Garbey at al. [9] detect the cardiac pulse at the distance from thermal face images, by using Fast Fourier Transform on features extracted from individual points along a line based region along major superficial vessels. Wavelet analysis is used by Fei and Pavlidis [8] to estimate the breathing information from the mean thermal signal of the nostril region. Wu et al. [16] propose Eulerian Video Magnification as a video oriented method based on spatial decomposition, temporal filtering and amplification to reveal hidden information that is impossible to see with the naked eye. The authors apply the method to visualize the flow of blood on the face area and also to amplify and reveal small motions.

ARCHITECTURE

To enable hand gesture interaction, the system needs to automatically detect the hands of the user. The 2D face detection is handled by an adapted Viola-&Jones object detector [14] and the hand contour detector is based on Active Appearance Models [7].

We implemented and integrated the models for face detection, 2D face and facial features shape extraction and heart rate detection, using C++ programming language, Boost::Thread library [2] for parallel computing and the open computer vision library OpenCV [12].

Face detection

Viola-&Jones features have been initially proposed as a fast method for performing object detection [15]. The hand detection uses four types of simple features derived from pixel intensities of video frame images. The first feature type is based on computing the difference between the sums of pixel intensities within two horizontally or vertically adjacent rectangular regions and represent edge indicators. Each of the features is effectively computed by making use of the integral image as an intermediate representation of the original video frame image. The integral image at location $(\boldsymbol{x}, \boldsymbol{y})$ is computed as sum of pixels contained in the rectangle defined from the top-left most pixel location to the current location $(\boldsymbol{x}, \boldsymbol{y})$. If $\boldsymbol{ii}(\boldsymbol{x}, \boldsymbol{y})$ is the integral image and $\boldsymbol{i}(\boldsymbol{x}, \boldsymbol{y})$ is the original image, then: $\sum_{x' \leq x, y' \leq y} \boldsymbol{i}(\boldsymbol{x}', \boldsymbol{y}')$. The hand classification is performed using discrete Adaboost, a supervised method that performs binary classification. A strong classifier is iteratively obtained by learning several weak classifiers that perform just better than random guessing. If added to the ensemble, each classifier is trained on a data set which is specially generated by applying resampling or reweighting on the instances of the original hand data set.

Face Segmentation

Active appearance model (AAM) - AAM [7] is a statistical method that handles shape and texture variations of photo-realistic appearance. Seen as a top-down approach, the AAM makes use of prior knowledge on the grey-level appearance, shape structures as well as their relationships, in order to build generative models for the global analysis of a specific class of objects.

The extraction of shape and texture of the face from an image is equivalent to an optimization problem that involves the criterion of minimizing the difference between the real face image and the one generated by the appearance model. The distance measure can be written as follows: $r(p) = g_{im} + g_m$, where g_{im} is the vector of grey level values of the face patch in the input image and g_m is the vector of grey level values for the face image as it is estimated by the current model parameters p.

The matching implies finding the optimal appearance model parameters which would lead to the minimization of a scalar measure on the image difference, such as the sum of squares of elements $E(p) = r^T r$. Building a face appearance model implies a prior acquisition of the face data set. Each face sample has to be consistently annotated with a set of landmark points. Each sample of face shape is described using n 2D landmark points by a shape vector: $\mathrm{x} = (x_1, \dots, x_n, y_1, \dots, y_n)^T$. An example of face shape annotation is illustrated in Figure 1. The face shape data set is formed by concatenating the face shape vectors x_i of all the face samples. Further statistical analysis is applied so as to obtain shape data samples represented in the same coordinate system.

The appearance model parameters c and shape transformation parameters t identify the position of the model points in the image. For matching, pixels are sampled around the image region g_{im} and then the texture model is computed $g_s = T_u^{-1}(g_{im})$. The model texture can be written as: $g_m = \overline{g} + Q_g c$ and so the difference between the model and the image may be written as follows: $r(p) = g_s + g_m$ and $p^T = (c^T, t^T, u^T)$. The appearance model may be written as: $x = \overline{x} + P_s W_s^{-1} P_{cs} c$, and $g = \overline{g} + P_g P_{cg} c$, where P_{cs} and P_{cg} are matrices describing the modes of variation and W_s is introduced as the weight that compensates for the units of distance and units of pixel intensity.

The fitting procedure is an iterative process that assumes for a specific current residual, a p is chosen so that the value of $|r(p + \delta p)|^2$ is minimum. The choice has the form: $\delta p = -Rr(p)$, where: $R = \left(\frac{\partial r^T}{\partial p} \frac{\partial r}{\partial p}\right)^{-1} \frac{\partial r^T}{\partial p}$. The term $\frac{\partial r}{\partial p}$ is computed once, during the training step, by altering the elements of p with some amount (typically up to 0.5 standard deviation) and by computing the residuals for the face images included in the training data set. The model face is fit to the image face during an iterative procedure that alters the model parameters c, pose t and texture transformation u until the error value fits into an acceptable range.

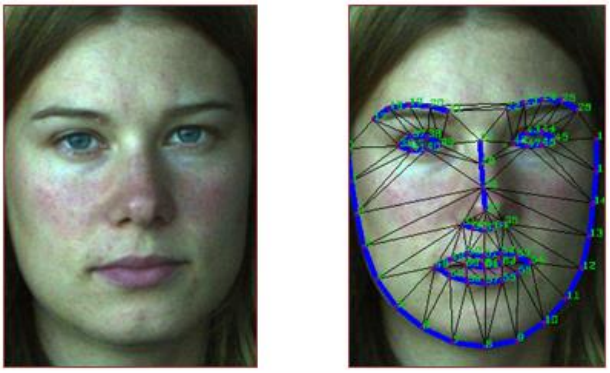

Figure 1: Active Appearance Model of face with 66 key points.

Heart Rate Analysis

Our heart rate detection algorithms are adapted versions of methods presented by Garbey et al. [9] and by Poh et al.[12][13]. At each step, the heart rate is computed based on the red, green and blue components of the face pixels from the video frames contained in the analysis window.

Each video frame is processed using the information from the Active Appearance Model to separate the red, green and blue channels of the face region (Figure 2) and of 10 facial feature ROIs (Figure 3). The values of each colour channel of a given ROI are averaged. As an intermediary step, a frame of the video sequence is represented by a vector of size three multiplied by the number of ROIs. For each ROI, the sequence of RGB averages is first applied a de-trending procedure. This step aims at removing the nonstationarities of the heart rate variability by removing the low frequency trend component of the signal [14].

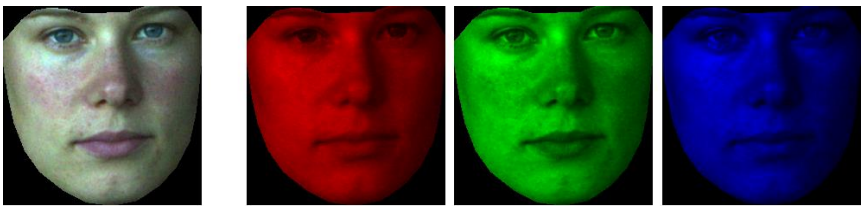

Figure 2: Color image to RGB channels decomposition for the whole face region - ROI.

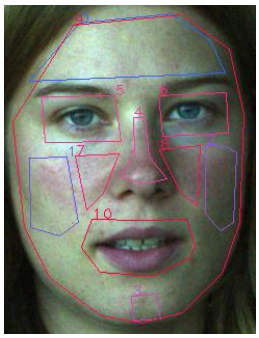

Figure 3: Facial feature region of interests - ROIs.

The parameters of the resulting high-pass filter are adjusted so that the de-trending procedure would not affect the range of frequencies of interest, between 0.7 Hz and 5 Hz (or between 42 bpm and 240 bpm).
Then the signals are normalized using: $y_i(t) = (x_i - \mu_i)/\sigma_i$, i=1, 2, 3, … , where μ_i and σ_i are the estimated mean and standard deviation of x_i. The left side of Figure 4 illustrates an example of smoothing and normalization (second row), given the initial RGB signals (first row). The smoothed and normalized RGB vectors are applied ICA that decomposes the signals into independent components (the right side of Figure 4). The signals are applied Fast Fourier Transform to obtain the power spectrum. Figure 5 shows the frequency magnitudes for the previous example of signals.
Although the order of the resulting independent components is not known, our tests indicated that the second channel contains the heart rate signal.
The analysis window contains two windows, one that corresponds to the history and one of the current data. Further on, at every step the analysis window is shifted with one second until the end of the signal is reached (Figure 6).

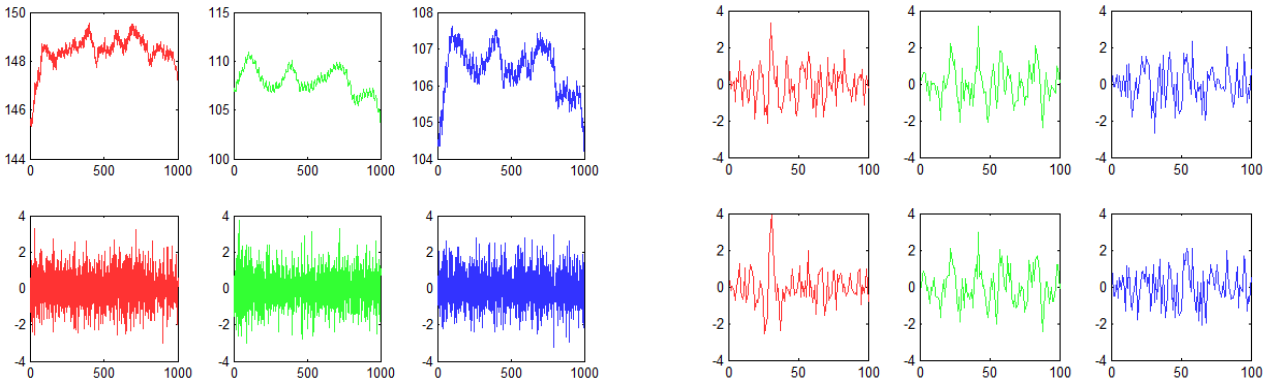

Figure 4: **Left**: first row: original RGB signals; second row: smoothed and normalized signals. **Right**: first row: de-trended and normalized signal of 100 samples; second row: ICA.

The current window is used to compute the current power spectrum. This is then used in a pointwise product with the historic power spectrum to obtain a filtered power spectrum of the signals. To remove the influence of accidental noise on the pulse signal, an extra step

assumes setting a threshold of 12 bmp to control the difference between two consecutive heart rate frequencies in the sense that no update is made if the difference exceeds the threshold value.

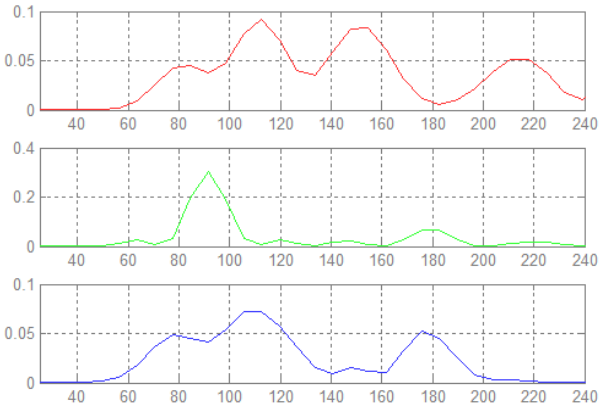

Figure 5: Frequency magnitudes for the interval [25,240] Hz, computed from 100 samples, at 30Hz.

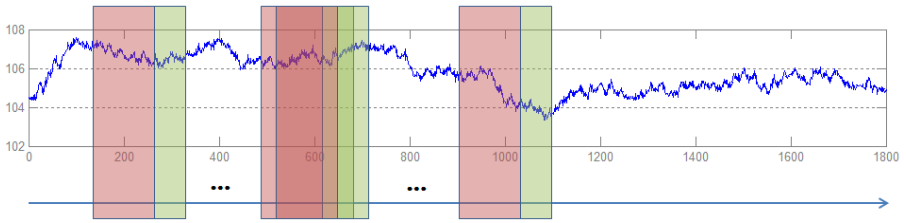

Figure 6: Fixed size window analysis.

EXPERIMENTAL SETUP

Prior to testing the models for automatic heart rate detection, we have prepared a small data set of video samples of different individuals. Six white persons participated at the recording, out of which one wearing glasses and another having beard. A Pike high speed camera from Allied Vision Technologies was used to take colour video recordings from frontal view and from profile. The subjects were asked to remain as still as possible for thirty seconds. One person was recorded twice, before and after making physical effort (climbing stairs). The parameters of Pike video camera were set as follows: shutter=100, gain=444, white balance=357, gamma=1 and pixel resolution of 252X350. This configuration of the camera parameters allowed for a runtime frequency of 242 frames per second. However, given the findings of our previous study according to which there is no actual gain in using video frequencies higher than 15fps, for testing the models we have resampled the original video to this frequency of 15fps. In addition to the video camera, the experimental setup included a device for accurate measurements of the pulse. We used Mobi 8 device produced by TMS International B.V., which is able to reliably measure the pulse at 2048 Hz. From the data acquired from the Mobi 8 we used only the pulse values measured at one second distance from one another. Each video obtained in the recording session had duration of 30 seconds and so was assigned a 30-elements vector as the pulse ground truth. The 30 elements vector is further applied moving averages of the order according to the size of the analysis window, including the history and current windows. For example, for a five second analysis window that is shifted with one second at every step of the algorithm we obtain 26 values following the application of the moving averages.

For testing, each video file was processed by the automatic system sequentially running the face detection component, face shape extraction component and the heart rate detection component. The rectangular region of interest ROI of the face computed by

the face detector is passed to the Active Appearance Model component that further on outputs the face and facial feature regions of interest. As part of the heart detection component, all pixels of each face ROI are scanned so as to extract the pixel RGB values. All these components were implemented using C++ programming language with OpenCV computer vision library [11]. The heart rate detection algorithm was implemented using MATLAB in combination with the Joint Approximate Diagonalization of Eigenmatrices - JADE, a MATLAB implementation for ICA [3].

Results

Figure 7 illustrates the root mean square – RMS of the heart rate errors as detected by two algorithms, given a range of values for the history window length (1-10 seconds) and a range of values for the analysis window length (1-15 seconds). The darker regions indicate better detection models with lower estimation errors. Figure 8 depicts the average processing times (in seconds) for detecting the heart rate, for each combination of history window length and current window length.

Figure 9 shows the box plot of the heart rate median errors and the 25th and 75th percentiles, for each face region separately and for the whole face ROI. The comparative study on face ROIs used a heart rate model for which the size of the current window is 2 seconds and the size of the history window is 4 seconds.

For both algorithms, the eye regions provide the lowest performance for detecting the heart rate (with errors of 40.86 bpm for the right eye and 37.93 bpm for the left eye in case of the first model and errors of 33.86 bpm for the right eye and 21.36 bpm for the left eye in case of the second model). The best heart rate detection for the first model is achieved for the left cheek outer ROI (error of 3.86 bpm) and for the second model for the left cheek inner ROI (error of 1.47 bpm).

The forehead ROI represents the second best region in terms of the pulse detection, giving an error of 4.14 bpm in case of the first model and an error of 1.49 bmp for the second model. The result of heart rate detection by taking the whole face ROI is worse than the best result with 58.88% for the first model (error of 5.85 bpm) and with 68.81% for the second model (error of 2.49 bpm).

If ROIs covering both sides of the face such as forehead, chin, nose and mouth are disregarded, left cheek and eye ROIs yield to better results than the right ROIs of the same face features. Moreover, the right cheek inner ROI performs better than the right cheek outer ROI. In case of the left cheek regions, the outer ROI performs better than the inner ROI in the case of the first model and the opposite in the case of the second model. The better results associated with the ROIs of the left face side can be explained by the fact that the participants to the tests had their right part of the face oriented inwards into the room and that skin pixel measurements could be slightly corrupted by the influence of distant fluorescent lamps mounted in the ceiling. We can therefore assume that in case there is no such light influence, the cheek and forehead regions provide the best pulse detection results.

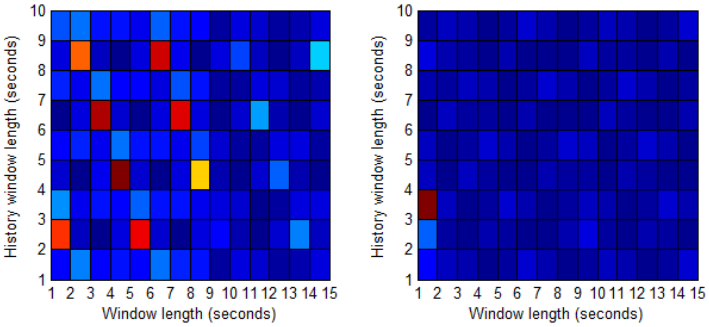

Figure 7: RMS test results for different sizes of the current and history windows.

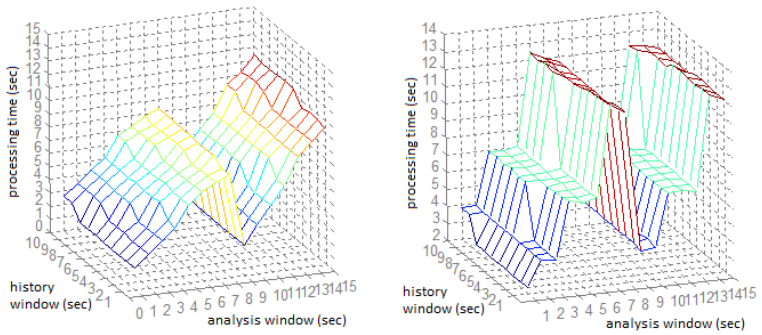

Figure 8: Processing time for different sizes of the current and history windows.

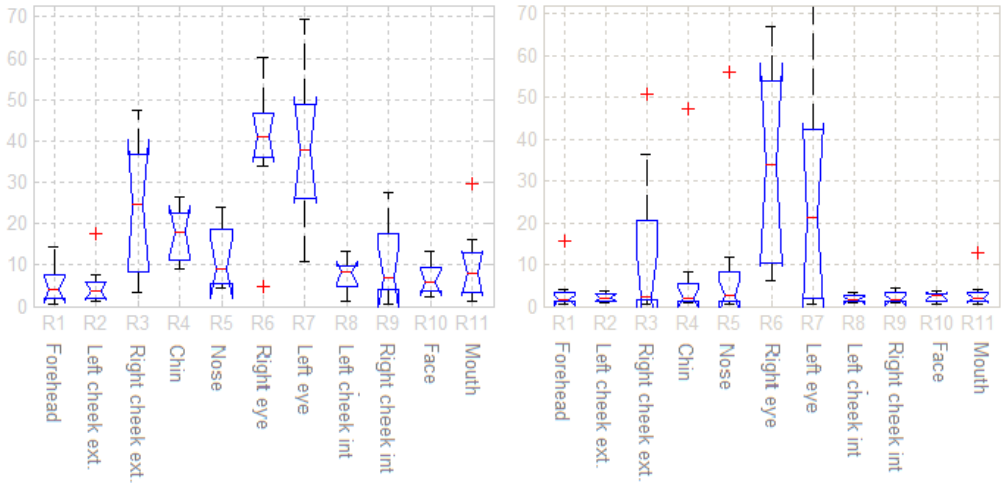

Figure 9: Box plots of heart rate median errors together with the 25th and 75th percentiles, for each face region.

CONCLUSIONS AND FUTURE WORK

The current paper presented a comparative study on the influence of face regions for heart rate detection from video sequences. Additionally a second novelty of our research is the integration of Active Appearance Models for accurate shape extraction of face and facial features. Following a set of tests, we clearly showed that the whole face area does not necessarily provide the optimal region of interest for pulse detection, the best being the cheek and forehead regions. The outcome of this research may be used to optimize the scanning of the face region which in turn leads to shorted computation times and more accurate results for the heart rate detection.

Next, we plan to extend the detection system with a method based on Doppler radar measurements. The goal is to obtain more robust heart rate analysis, running even in conditions of real life head motion. These would make contactless heart rate analysis available for real-time remote monitoring of people acting naturally during conversation, while working and playing.

REFERENCES

[1] Allen, J. Photoplethysmography and its application in clinical physiological measurement. In Physiological Measurement, vol. 28, pp. R1-R39, 2007.

[2] Boost::Thread, http://www.boost.org/doc/libs/1_52_0/doc/html/thread.html.

[3] Cardoso, J.-F. and Souloumiac, A. Blind beamforming for non-gaussian signals. In IEE Proceedings-F, vol. 140, no. 6, 362–370, 1993.

[4] Cardoso, J.-F. High-order contrasts for independent component analysis. In Neural Comput., vol. 11, pp. 157–192. 1999.

[5] Cidota, M. A., D. Datcu, L. J. M. Rothkrantz. Learning AAM fitting with kernel methods, 11th International Conference on Artificial Intelligence and Soft Computing - ICAISC, Springer Lecture Notes in Artificial Intelligence, Zakopane, Poland, May, 2012.

[6] Comon, P. Independent component analysis, a new concept? In Signal Processing, 36, 287-314, 1994.

[7] Edwards, G. J., C. J. Taylor, T. F. Cootes. Interpreting face images using active appearance models. Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. p. 300, 1998.

[8] Fei, J., I. Pavlidis. Thermistor at a Distance: Unobtrusive Measurement of Breathing. IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, 988-998, 2010.

[9] Garbey, M., N. Sun, A. Merla, I. Pavlidis. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. In IEEE Trans.Biomed. Eng., vol. 54, no. 8, 1418 –1426, 2007.

[10] Greneker, E. Radar sensing of heartbeat and respiration at a distance with application of the technology. In RADAR, vol. 97, no 449, 150–154, 1997.

[11] OpenCV computer vision library, http://opencv.willowgarage.com/wiki/.

[12] Poh, M.-Z., D. J. McDuff, R. W. Picard. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam, IEEE Transactions on Biomedical Engineering, vol. 58, no. 1, pp.7-11, 2011.

[13] Poh, M., D. McDuff, R. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. In OPTICS EXPRESS, vol. 18, no. 10, 2010.

[14] Tarvainen, M., P. Ranta-Aho, P. Karjalainen. An advanced detrending method with application to hrv analysis. In IEEE Trans. Biomed. Eng., vol. 49, no. 2, 172 –175, 2002.

[15] Viola, P., M. Jones. Robust Real-time Object Detection, International Journal of Computer Vision, 2001.

[16] Wu, Hao-Yu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W. T. Freeman. Eulerian Video Magnification for Revealing Subtle Changes in the World ACM Transactions Graph. (Proceedings SIGGRAPH'12), vol. 31, no.4, 2012.

ABOUT THE AUTHORS

Ing. Dragos Datcu, PhD, PostDoc., Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628BX Delft, The Netherlands, E-mail: D.Datcu@tudelft.nl.

Assistant Prof. Marina Cidota, PhD, Faculty of Mathematics and Computer Science, University of Bucharest, Academiei Street, Bucharest, Romania, E-mail: cidota@fmi.unibuc.ro.

Associate Prof. Stephan Lukosch, PhD, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628BX Delft, The Netherlands, E-mail: S.Luckosch@TUDelft.nl.

Prof. Leon Rothkrantz, PhD, Sensor Technology, SEWACO Department, Netherlands Defence Academy, Nieuwe Diep 8, 1781 AC, Den Helder, The Netherlands, E-mail: L.J.M.Rothkrantz@TUDelft.nl.