A concept of bimodal visual emotion recognition in computer users

Jaromir Przybyło, Eliasz Kańtoch, Piotr Augustyniak

AGH University of Science and Technology, 30 Mickiewicza Ave. 30-059 Kraków
Poland, przybylo@agh.edu.pl

Abstract. Touchless measurement of affects in computer users is gaining much interests in current man-machine interactions. In this work we present the concept of bimodal visual emotion recognition in computer users. Our idea builds on two different paradigms: a pulse detection based on face image processing and an analysis of scanpath features using eyetracking. The concept is supported by mutual correspondence of the two different methods, while both originate from a video frame sequence possibly acquired with a single sensor. Besides the novel concept, we also put forward several suggestions for future work.

Keywords: Emotion recognition, videopletysmography, scanpath analysis, eyetracking

1 Introduction

Computer users' emotions are currently considered as an effective mean to improve man-machine interactions, avoid misleading presentation of information, controlling of virtual personalities (bots), real assisted living systems or games. Recent development in biosignal sensing technologies allows to monitor human vital signs during everyday tasks such as doing computer work or playing computer games. The emotional state of the human is correlated with the heart rate measured by various methods including electrocardiography or infra-red plethysmography.

Most of systems presented as far require a wearable device which may be cumbersome or intrusive in some scenarios. The novelty of our approach consists in a touchless emotion recognition based on two complementary paradigms and visual systems. The touchless measurement of affects in computer users is gaining much interests in future man-machine interactions.

In this work we present the concept of bimodal visual emotion recognition in computer users. The remaining part of this paper is organized as follows. Section 2 describes the material and methods. Section 3 presents principal findings and discussion of the results and concludes the paper.

2 Material and methods

The proposed concept is based on image analysis and two different paradigms of emotion recognition previously studied by our group: a pulse detection and a scanpath features analysis.

2.1 Hardware set-up

The hardware set-up included: a computer, a web camera, the wireless ECG recorder Aspekt 500, the wearable heart rate monitor (two-electrode chest sensor H7 by Polar) and the eyetracking multisensor measurement system JAZZ-novo. The experiment set-up allowed simultaneous acquisition of the electrocardiogram and eye movements together with the operator's face video signal. The computer was running a video acquisition software, a custom-developed ECG acquisition software and MATLAB scripts for control of the experiment protocol. The camera built-in microphone redundantly recorded the ambient audio signal (16 kHz sampling frequency) in order to synchronize recordings to the auditory stimulus. Volunteers were recruited in the study. They were asked to wear the measurement devices while working on the computer or playing. The hardware set-up diagram is shown in Fig. 1. and a sample volunteer during the experiments is presented in the Fig. 2.

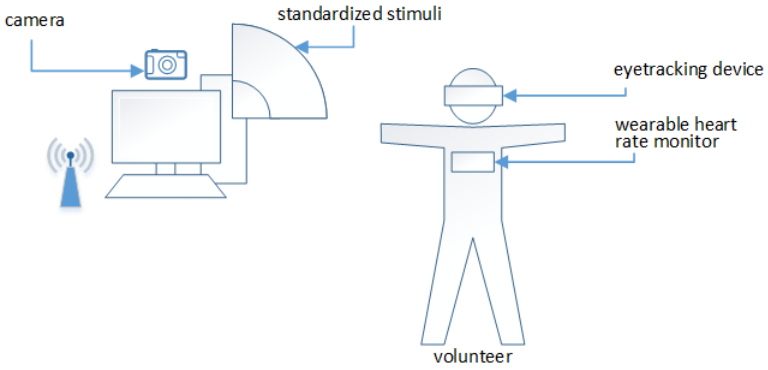

Fig. 1. The hardware set-up diagram

2.2 Face detection method

Face detection is a common procedure required by each of the two paradigms prior to emotion detection. There are many sources of variation in facial appearance. They can be categorized [3] into two groups - intrinsic factors related to the physical nature of the face (identity, age, sex, facial expression) and extrinsic

Fig. 2. The volunteer during experiments

factors resulting from the scene and lighting conditions (illumination, viewing geometry, imaging process, occlusion, shadowing). All these factors make face detection and recognition (especially distinguishing various facial actions) a difficult task. Therefore, many approaches to face detection and recognition in natural conditions appeared in recent years. A surveys of those methods are presented in articles [18], [17]. One such well-known approach to face detection that has the most impact is the work by Viola and Jones [16]. The proposed face detector can run in real time and it is based on the integral image, classifier learning with AdaBoost and the attentional cascade structure. Also, there are many high quality publicly available code repositories with efficient implementations of this algorithm, for example the OpenCV library [8]. Another fast and reliable face detection method can be found in DLib library [2]. It is based on the Histograms of Oriented Gradients (HoG) algorithm proposed in [1], combined with Max-Margin Object Detection (MMOD) [6] which produces high quality detectors from relatively small amounts of training data. Wide availability of databases and benchmarks for 'in-the-wild' object detection, stimulated research on face detection and recognition using Deep Convolutional Neural Networks DCNN [5] [9]. One of the advantages of the DCNN is that it can be used either for feature extraction or/and classification tasks, thus eliminating the need of finding good discriminative face features. Video-based heart rate estimation requires the face to be detected continuously i.e. in each of the video frames. In case when the face cannot be detected, the KLT tracking algorithm [14] can be used to estimate face position in the consecutive frames (Fig. 3).

In present work we combine Dlib face detector (which seems to be faster and

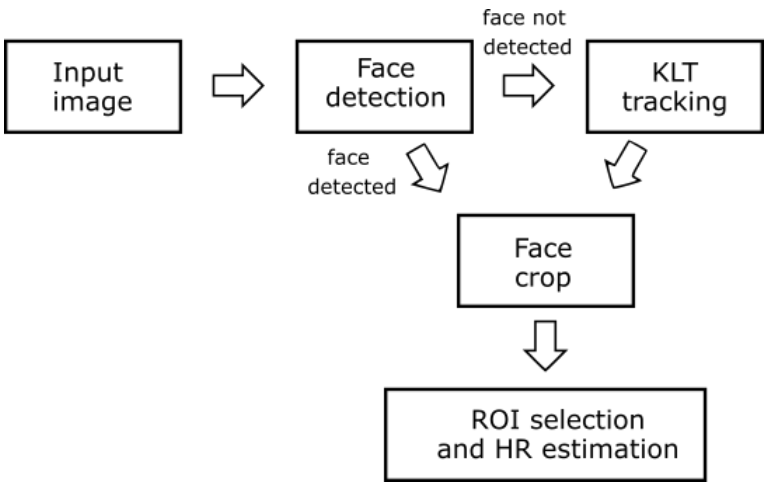

Fig. 3. Face detection and tracking algorithm

more robust than the Viola-Jones detector) with the KLT tracking algorithm to efficiently track the face in the video sequence. In each frame we try to detect and localize the face. If the face is not detected, the KLT tracker is used to track facial features and to estimate its position between frames. Then a Region of Interest (ROI) is computed based on the object detection and tracking results. It is used to calculate the mean value of local color components inside the ROI. These values are stored in a buffer of the length of N points for the heart rate estimation.

In case of computer users, some geometry-related assumptions facilitate the face recognition [11]. Provided the user is facing the screen, which frequently is the most natural position, the camera mounted on the top of the screen captures the face of the user. Moreover, the computer is usually operated by a single user, thus the detection of multiple faces is not applicable.

2.3 Visual detection of pulse

Videoplethysmography initiated by the paper by Verkruysse [15] is currently recognized as a promising noninvasive heart rate measurement method, advantageous for ubiquitous monitoring of humans in their natural living conditions [10], [7]. Using this method in an unstable lighting conditions (e.g. in computer users with faces illuminated by the screen) was studied in our previous paper [11]. We examined various image acquisition aspects including the lighting spectrum, frame rate and compression and their influence on pulse detection accuracy (Fig. 4). We implemented a pulse detection algorithm based on the power spectral density, estimated using Welch's technique. The experimental results showed that lighting conditions and selected video camera settings, such as compression and the sampling frequency, influence the heart rate detection accuracy, however

in a stable light from a fluorescent source the heart rate measurement error was as low as 0.35 beats per minute (bpm).

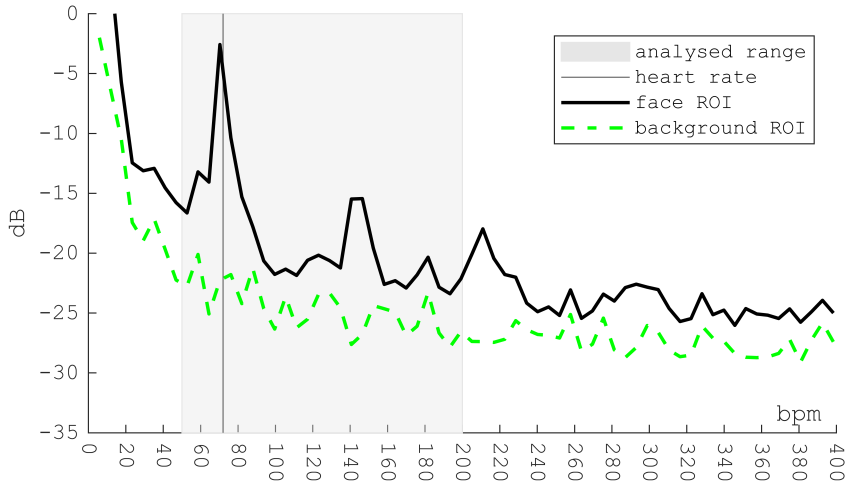

Fig. 4. Example power spectrum density of a video sequence recorded with artificial fluorescent light, 200 FPS.

In the particular case of computer users there are two methods for compensation of the illumination changes:

- the *á priori* analysis of the displayed content, and
- the *á posteriori* analysis of the features of acquired image.

The analysis of the displayed content provides usable information to the compensation algorithm about the average color of the screen, possible scene changes or flickering (Fig. 5). This method does not take into account external illumination sources and thus has to be complemented by the *á posteriori* analysis of the features of acquired image. The latter method, however, operates on captured visual scene and may cause delays impairing the real-time detection of emotion. In [4] authors describe a color-based face segmentation algorithm with automatic calibration of skin-color model parameters, using a flash illumination from the computer screen. The proposed algorithm allows to increase the efficiency of face segmentation, and can be used to *á priori* analysis of the displayed content.

2.4 Emotion-related scanpaths changes

In [13] we proved the emotional influence on the human scanpath. This method assumes that the emotional state of the observer can distract his or her visual attention and can be reliably expressed in parameters of eye movements. We

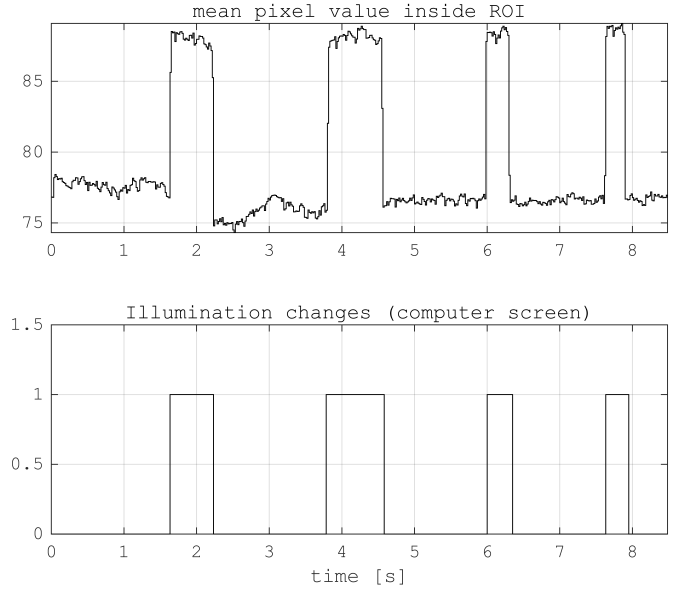

Fig. 5. Changes of the intensity of pixel values inside the face ROI due to variation of face illumination by the computer screen

performed visual task experiments in order to record scanpaths from volunteers under the auditory stress of controlled intensity. We used a calibrated set of stimuli and recorded the response of volunteers' central nervous system expressed by the heart rate. For each efficiently stimulated participant, we related the scanpath parameters to the accuracy of solving a given task and to participant's comment. As a principal result, we obtained a significant change of:

- the saccade frequency in 90 % and
- the maximum velocity in fixation phase in 80 %

of participants under stress. These results prove the influence of emotions on visual acuity, feasibility of eyetracking-based assessment of emotional state and motivate further investigations. In our studies, as other sources suggest, we employed a thorough eyetracker calibration procedure. In case of computer users, the calibration can be simplified to a procedure that consists of following a point, video or other graphical element that moves across the screen. Moreover, there are commercially available low-cost eyetrackers designed for PC gaming and laptop users.

Otherwise, in case of computer users, the calibration can eventually be omitted. In several everyday usage scenarios computer users interact with the content immediately as it is displayed, thus focus their gaze accordingly to newly displayed items. This concept is worth further studies on whether temporal syn-

chronicity of the displayed content (e.g. a text prompt) provides accurate data to substitute geometrical calibration of the gaze focus and the displayed cue. Additionally, the method assumes the geometrical position of the eyetracking camera to be constant with respect to the displaying screen. It is then worth studying whether the face detection, performed independently on gaze tracking, yields sufficient data for rough positioning of the eyes. Finally, possible errors in gaze tracking may not be significantly influencing the accuracy of calculation of saccade frequency or maximum velocity in fixation phase.

2.5 Combining data from two visual measurement modes

At first glance the simultaneous usage of two independent information sources improves the correctness of emotion recognition, however optimal use of bimodal data needs more investigation. Current results show high degree of correlation of the heart and eyetracking parameters. However, the investigations were made in stationary conditions where the stimulus was absent or present, while a static stimulus rarely occurs in real life scenarios. Little is known about the dynamic responses of cardiovascular and perceptual systems to the sudden presence of the stimulus. Another studies should consider the possible habituation to stimuli of long duration.

On the other hand, auditory stimuli applied in our experiments are standardized in two dimensions: intensity and valence. Further experiments should clarify whether both responses are equally related to the stimulus parameters or there is another nonlinear model that optimally maps the 2D stimulus space to the 2D cardio-perceptual response.

3 Discussion and conclusions

In our work we showed the high degree of the correlation of the heart and eyetracking parameters. However, the study had some limitations including laboratory conditions. The presented bimodal approach can be used to increase the accuracy of existing methods of emotion recognition based on the analysis of facial expressions. It may help to provide more data about subject's vital signs and as a result - to facilitate the emotion recognition without the need of sophisticated wearable sensor systems. The advantage of our approach is twofold:

- it allows for a touchless measurement, with application of suitable optics even a distant measurement is possible, and
- provided the face recognition is accurate, it enables a calibration-free operation, and therefore could be used confidentially (i.e. in an unconscious computer operator).

Although two different paradigms were proposed for a concurrent and complementary use, it doesn't necessarily require two different sensors. Our future work will pave the way for using a single high resolution camera (preferentially mounted above of the laptop screen) to provide image for both eye motion and

pulse wave detection. Several other issues we should address in our research plans include:

- Facial action recognition could also provide additional insight into user's emotional state. However, measuring involuntary emotions from video sequences is still a challenge.
- Face detection and tracking can be extended to eye-gaze estimation. However, further study is required to verify if gaze tracking accuracy impacts the calculation of saccade frequency or maximum velocity in fixation phase.
- The concept of using temporal synchronicity of the displayed content (e.g. a text prompt) to calibrate eye-tracking requires further studies.

4 Acknowledgment

This scientific work is supported by the AGH University of Science and Technology in year 2018 as a research project No. 11.11.120.612.

References

1. Dalal, N., Triggs, B. Histograms of oriented gradients for human detection. In CVPR 2005, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 2005, pp. 886-893.
2. http://dlib.net/
3. Gong, S., McKenna, S.J., Psarrou, A.: Dynamic Vision, From Images to Face Recognition. Imperial College Press, London 2000.
4. Jabłoński, M., Przybyło, J., Wołoszyn, P. Automatic face segmentation for human-computer interface (in Polish) Automatyka, 9, 2005, 587-600.
5. Jiang, H., Learned-Miller, E. Face detection with the faster R-CNN. In 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, 650-657.
6. King, D.E. Max-margin object detection. arXiv preprint, 2015, arXiv:1502.00046.
7. Li, X., Chen, J., Zhao, G., Pietikainen, M. Remote heart rate measurement from face videos under realistic situations. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, 4264-4271.
8. https://opencv.org/
9. Parkhi, O. M., Vedaldi, A., Zisserman, A. Deep Face Recognition. In BMVC Vol. 1, No. 3, 2015, p. 6.
10. Poh, M.Z., McDuff, D.J., Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics express, 18(10), 10762-10774.
11. Przybyło J.: Vision Based Facial Action Recognition System for People with Disabilities. Proc. 8-th international conference on Human System Interactions (HSI2015), 2015, 244-248
12. Przybyło, J., Kańtoch, E., Jabłoński, M., Augustyniak P. Distant Measurement of Plethysmographic Signal in Various Lighting Conditions Using Configurable Frame-Rate Camera. Metrol. Meas. Syst., Vol. 23, 2016, No. 4, pp. 579-592

13. Przybyło, J., Kańtoch, E., Augustyniak, P. Eyetracking-based assessment of affect-related decay of human performance in visual tasks. Future Generation Computer Systems, 2018, https://doi.org/10.1016/j.future.2018.02.012
14. Tomasi, C., Kanade, T. Detection and tracking of point features. International Journal of Computer Vision 1991.
15. Verkruysse, W., Svaasand, L.O., Nelson, J.S. Remote plethysmographic imaging using ambient light. Optics express, 16(26), 2008, 21434-21445
16. Viola, P., Jones, M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, pp. I-511-I-518
17. Zafeiriou, S., Zhang, C., Zhang, Z. A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding, 138, 2015, 1-24.
18. Zhang, C., Zhang, Z. A survey of recent advances in face detection. 2010