Emotion Recognition - the need for a complete analysis of the phenomenon of expression formation

Katarzyna Bobkowska[1], *Marek* Przyborski[1,*], and *Dariusz* Skorupka[2]

[1]Gdansk University of Technology, Geodesy Department, 80-233 Gdansk, Gabriela Narutowicza St. 11/12, Poland
[2]General Tadeusz Kościuszko Military Academy of Land Forces, Production Department, 51-147 Wroclaw, Czajkowskiego St. 109, Poland

Abstract. This article shows how complex emotions are. This has been proven by the analysis of the changes that occur on the face. The authors present the problem of image analysis for the purpose of identifying emotions. In addition, they point out the importance of recording the phenomenon of the development of emotions on the human face with the use of high-speed cameras, which allows the detection of micro expression. The work that was prepared for this article was based on analyzing the parallax pair correlation coefficients for specific faces. In the article authors proposed to divide the facial image into 8 characteristic segments. With this approach, it was confirmed that at different moments of emotion the pace of expression and the maximum change characteristic of a particular emotion, for each part of the face is different.

1 Introduction

The problem of recognizing emotion based on image analysis is known worldwide. There are many algorithms created by specialists from different fields that allow such analyzes. Examples of such algorithms and analysis are described in the papers [1], [2]. Unfortunately, each of these tools is not reliable. Standard algorithms allow to distinguish only a few basic emotions. It is important to know, there are dozens of emotional states for which there is no algorithm for remotely recognizing these states. To create a reliable tool for remote sensing of emotions, consciousness of the action of the human organism is essential. The big problem is that human facial expressions can be consciously controlled by ourselves [3]. Therefore, the authors consider that observation with high accuracy is important (this translates into high speed of registration) the whole phenomenon of emotion from its beginning (excitement stimulus) to the end. In addition, the analysis of the change occurring on the face should be done without generalizing to one image showing expression. Presented approach will allow you to record all facial changes, including very short-lasting, so-called microexpressions. They are not noticeable by people observing this phenomenon without the use of additional tools, however our brain is able to understand them even without focussing attention. This is caused by the time in which microexpression on the face is visible. Usually there are hundredths of a second. Recent years of research on microexpressions have evolved at a significant pace. It is believed that recognized microexpressions are the key in the process of knowing the true feelings you feel, for example, in people who want to hide such emotions (criminals, terrorists) or in people with mental illness (eg schizophrenia). Research conducted by scientists from various fields shows promising results to detect "hidden emotions" [4-6]. The work of deepening knowledge on this subject can help to create a reliable tool to explore what a person is trying to hide and unconsciously shows on his or her face.

2 Acquisition and processing of data

The authors, have developed an experiment that was based on stimulating emotions (random anger or joy). The stimulus that triggered the emotional state was the randomly selected images from the Karolinska Directed Emotional Faces (KDEF) [7]. 70 people were examined (30 cases were analyzed for this article). A high speed camera was used to register video material. The recording was done at a rate of 1000 frames per second, in grey scale, in high resolution. Each film contained frames representing the face during:
• neutral state,
• stimulation of emotions,
• expression during the emotion
• expulsion of expression.
The resulting videos occupied approximately 6 GB of data on average.
In the study of emotion analysis, high speed cameras are rarely used. Analyzes are made on single images or a stack of images obtained with a standard speed camera. In [8] work, attention has been focused on the need for microexpression analysis with non-standard registration speed. The authors of this work have used rapid

---

* Corresponding author: marek.przyborski@pg.edu.pl

recording cameras at 100 fps and 200 fps. The study data at similar recording speed were collected by the authors of the paper [9]. It can be said that for more accurate measurements and analyzes it is necessary to apply a non-standard registration rate. This has also been confirmed by the authors of this publication in the following articles [10-12].

The initial process of collecting research material and preparing for further analyzes is shown in Fig. 1. Each film was divided into single frames in .tiff format and stored in 8-bit depth as an A_i image - where i stands for the number of frames in the film.

Fig. 2 shows an example of the distribution of a film into individual frames. It was important to normalize the entire research material by stretching the histogram. Therefore it is possible to obtain equally qualitative data. In such a way the prepared images could be subjected to further analyzes.
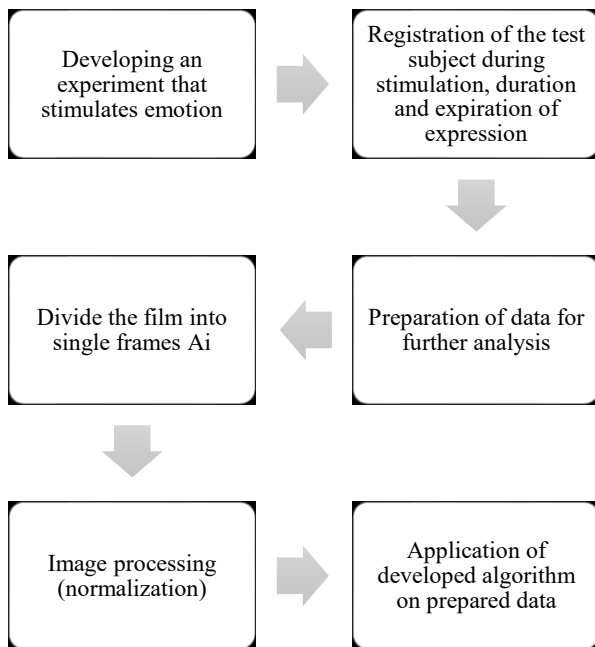
Fig. 1. Registration and preparation of data.

The next step was to develop an algorithm. Fig. 3 shows the flow chart. A completely new approach has been set up. This approach assumes the use of known methods: face detection, mouth detection, eye detection [13], digital image correlation. The initial idea is to divide the face image into specific segments. To achieve this, at the very beginning from the stack of images representing the face with the background (A_i images), the image representing only the face (C_i images) should be cut out. Each of these images for the same movie should have the same size, because the correlation coefficient can only be determined for images of the same size (resolution). Assuming that the image is a matrix, within a single movie, the condition must be fulfilled:
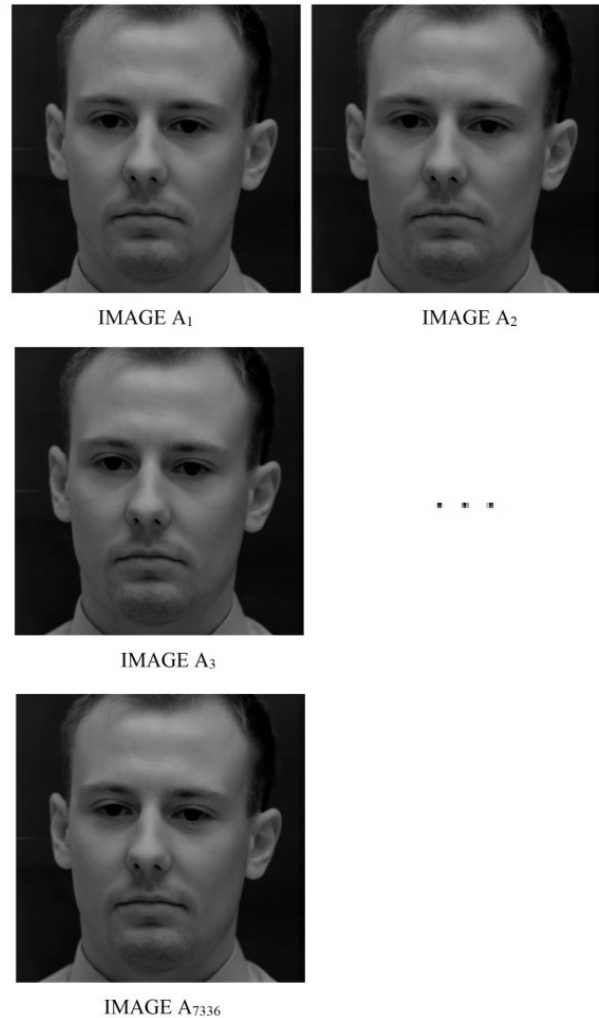
For $C_{1(mn)}$ and $C_{i(gh)}$

$$m = g$$

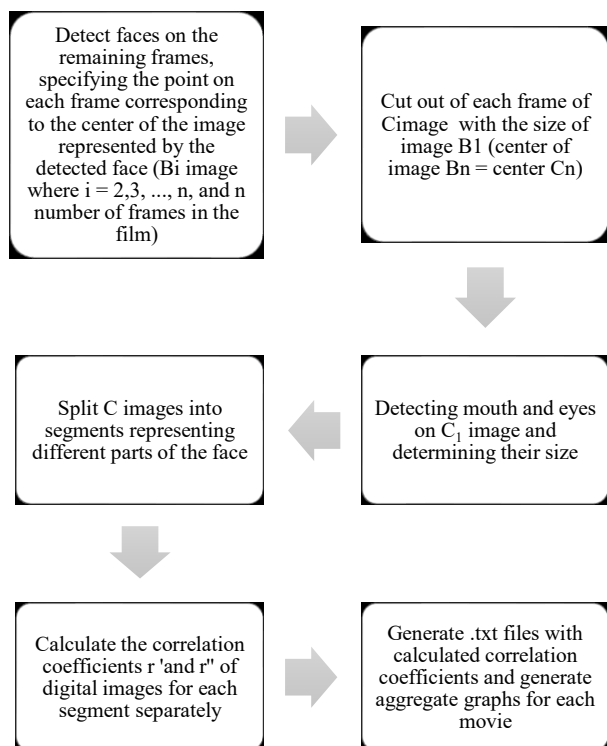$$n = h$$

where:
m and g – number of matrix columns,
n and h - number of matrix rows.

Fig. 2. The effect of unfolding the film on single frames on the example of a film consisting of 7336 frames.

Therefore, the created algorithm assumes the cut, from the original images, the images of a dimension such as the assumed image (A_i) for the first frame. However, the most important aspect was that in every image, the face should be centered. This was achieved by detecting a B_i image representing only the face on all images, while the matrix dimensions of these images were not always the same. Therefore, a dependency has been defined, which allowed to cut images satisfying this condition.

The next stage was eye and mouth detection (Fig. 4 and Fig. 5). This allowed us to create a framework for dividing the frame into eight segments. In this case, the image of the eyes and mouth had to be of the same size as their counterparts for each of the cages. Therefore, for all C_i images, an division was proposed based on the C_l image identical to detect these parts of the face using existing available functions of Matlab.

Detect faces on the remaining frames, specifying the point on each frame corresponding to the center of the image represented by the detected face (Bi image where i = 2,3, ..., n, and n number of frames in the film)	Cut out of each frame of Cimage with the size of image B1 (center of image Bn = center Cn)
Split C images into segments representing different parts of the face	Detecting mouth and eyes on C_1 image and determining their size
Calculate the correlation coefficients r 'and r'' of digital images for each segment separately	Generate .txt files with calculated correlation coefficients and generate aggregate graphs for each movie

Fig. 3. Description of the procedure used in the algorithm.

Fig. 4. Detected eyes from picture C_1.

Fig. 5. Detected mouth from picture C_1.

On this basis, an automatic pattern for dividing the facial image into eight segments was created, as shown in Fig. 6. The segments were represented as:
• Segment 1 - lips with a beard
• Segment 2 - lower right cheek,
• Segment 3 - bottom left cheek,
• Segment 4 - eyes,
• Segment 5 - nose,
• Segment 6 - upper right cheek,
• Segment 7 - upper left cheek,
• Segment 8 - forehead.

Fig. 6. Scheme of image C1 division.

The next step was to perform the calculation of the correlation coefficient r' and r''. The dependencies and formulas have been thoroughly described in other author's publications [10], [11]. The whole process was automatic, using the author's algorithms.

The results generated by the application were used for further analyzes to investigate information about emotion-induced facial changes. Among other things, the speed of the changes taking place on a particular segment was analyzed, the comparison of the attainment of the correlation coefficient for each segment, and the intensity of the change.

3 Results

As a result of calculations have been created matrices of one row containing values of correlation r' (for example $R'_{S1} = [0.994 \quad 0.991 \quad 0.982 \quad ... \quad 0.976]$, where S1 - segment no 1) and graphs showing changes (Fig. 7) and r'' coefficients for each segment of the test face. The graphs represent a change in the correlation coefficient depending on successive images.
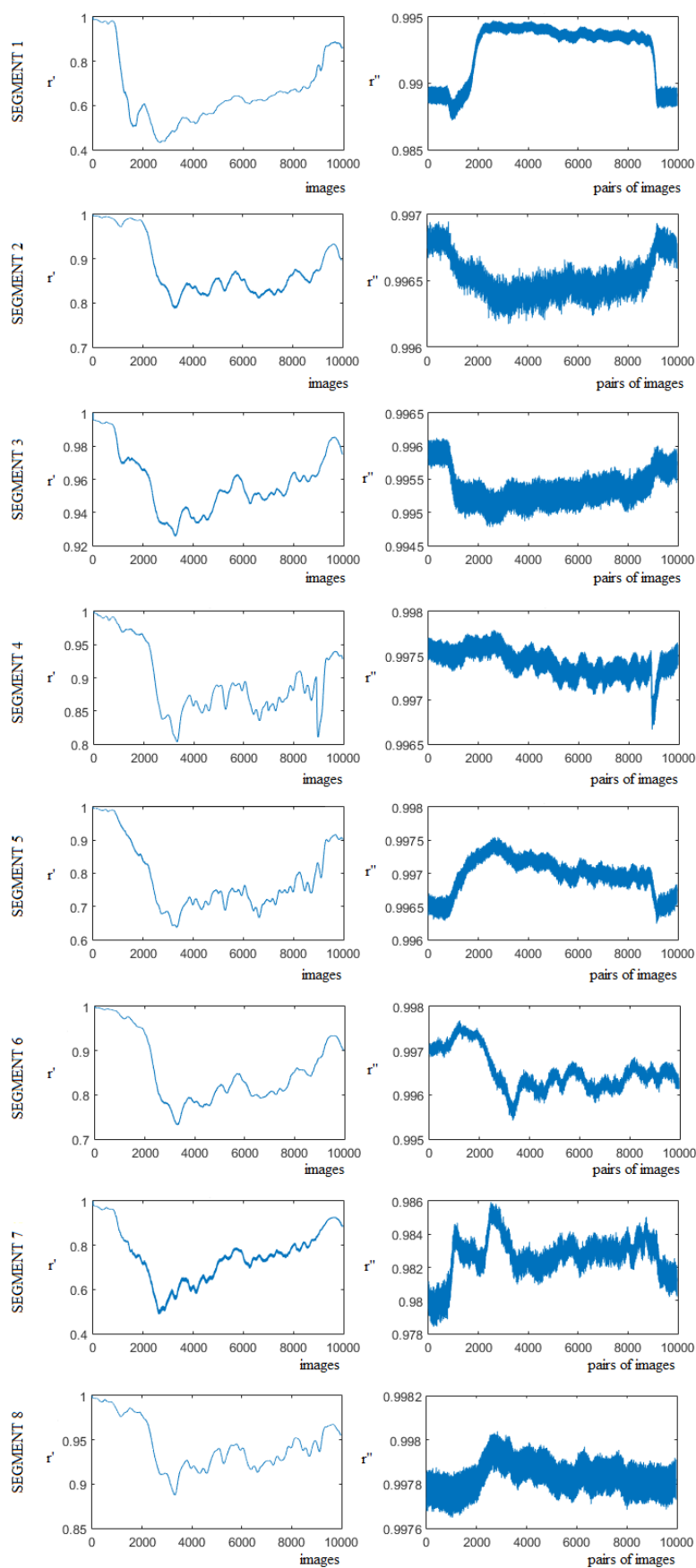
Fig. 7. Variability of correlation coefficient for individual segments (emotion - joy).
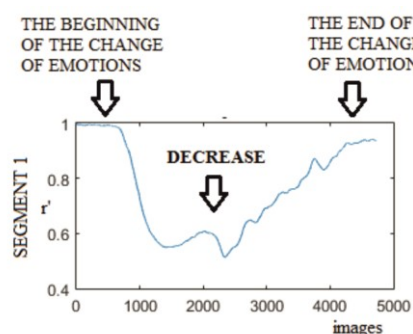
Fig. 8. Interpretation of the graph of the correlation coefficient r' between the first image and sequences (segment 1, emotion - joy).
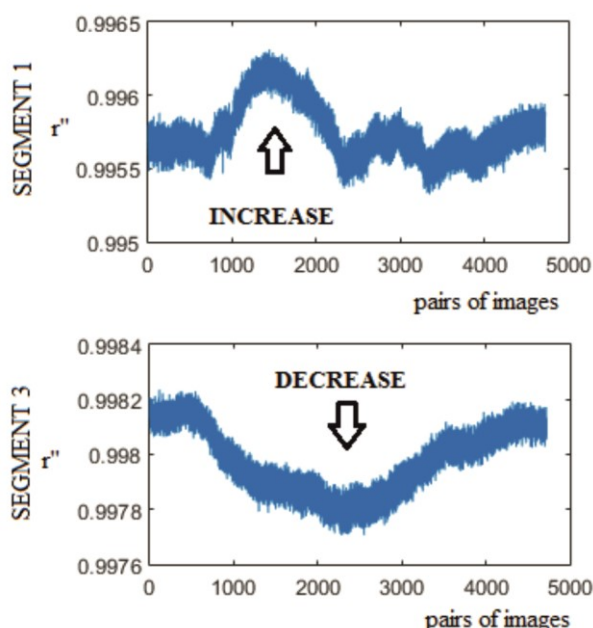
Fig. 9. An example of the increase and decrease of the correlation coefficient between adjacent pairs of images for Segment 1 and Segment 3 (emotion – joy).

Taking into account only the obtained graphics, simple relationships can be noticed. The beginning of change of emotional state and its termination is visualized (Fig. 8), by decreasing the coefficients r' for each segment. The graphs showing the change of the correlation coefficient r" are different. At the moment of emotion, the drop in coefficient is not always noticeable. In some cases, growth is noted (Fig. 9). The article did not undertake a thorough analysis of data r". The results are promising and will be used in future research.

3.1 Detection of batting an eye

For all cases, a sudden increase or decrease in the graph showing the correlation coefficient r" for a given segment represent a rapid change in this area. For example, the files on the graphs generated by the algorithm for segment 4 (representing the eyes of the examined person) show a blink of an eye.

3.2 Comparison of reaching the threshold of decrease in correlation coefficient

By analysing all the graphs for one test person, it is noted that for each segment, the correlation coefficient starts to decrease somewhere else. Based on this, it can be said that not all parts of the face react at the same time and with the same intensity at the stimulus.

Already at the "first glance" the joy of the other person is noticeable, after his smile, which is characteristic of this emotion. And the same is true in case of anger. We recognize it in other people after a specific look. The analyser show that it is not right to focus solely on one of all parts of face. In some cases, before there is a smile on the face representing joy, there are first changes in other areas. To provide this, the following calculations were made:

$$a_k = \overline{R'_{Sk}} \tag{1}$$

$$b_k = \frac{\max(R'_{Sk}) - \min(R'_{Sk})}{2} \tag{2}$$

$$c_k = \max(R'_{Sk}) - (\max(R'_{Sk}) - \min(R'_{Sk}) \cdot 0{,}25) \tag{3}$$

where:

k- segment number

$\overline{R'_{Sk}}$ - mean of the correlation coefficients stored in the matrix R'_{Sk}

In next step, with the command, the number of pairs of images for which values lower than a_k, b_k, c_k were appointed, The results are presented in Tables 1, 2, 3 for example persons. In this way, how quickly responded parts of the face to a given stimulus can be compared.

Table 1. Obtained values a_k.

Test	Segment no k							
	1	2	3	4	5	6	7	8
16	**1236**	2468	2285	2440	2314	2364	1636	2422
28	780	630	645	730	687	**533**	536	712
51	**970**	1196	993	2065	1985	1480	1314	1802
52	1409	1407	1395	781	1484	**586**	1458	1356
22	1644	1232	1407	1303	1228	1211	1315	**940**
24	952	1057	888	1155	885	2651	**775**	2005
59	851	1013	**845**	861	911	915	881	934

Table 2. Obtained values b_k.

Test	Segment no k							
	1	2	3	4	5	6	7	8
16	**1163**	2414	2157	2420	2217	2328	1695	2420
28	809	540	664	793	815	**511**	846	783
51	**974**	2060	1021	2184	2076	2174	2059	2055
52	1396	1407	1405	791	1475	**2**	1416	887
22	4512	1392	1496	1392	1292	1389	1496	**920**
24	868	1077	747	1254	827	961	774	2037
59	**811**	818	775	826	861	841	838	900

Table 3. Obtained values c_k.

Test	Segment no k

Test	1	2	3	4	5	6	7	8
16	1003	2227	**964**	2186	1416	2056	1024	2159
28	702	**395**	446	529	646	403	494	513
51	851	1040	**839**	2017	1236	1169	1239	1122
52	1120	1087	1325	585	773	**2**	828	397
22	1286	1112	1162	929	1091	1122	1266	**826**
24	756	728	**427**	1027	754	697	677	1155
59	755	**741**	746	769	801	761	793	815

3.3 Achieving the maximum for the segment

In addition to the fact that at different moments of emotion, the rate of change and the maximum change that presents emotion for each part of the face is different, is selection the following values for the test person - pairs of images representing R'_{Sk}. This dependency for several people is shown in Table 4. The values shown in the table confirm that the maximum facial changes do not occur at the same time for the entire face. Therefore, the analysis of a single photo for the purpose of recognizing emotions is unreasonable because it does not give full information about the expression.

Table 4. Obtained values R'_{Sk}.

Test	1	2	3	4	5	6	7	8
16	2673	3268	3268	3324	3337	3318	**2617**	3308
28	961	938	2922	930	944	2922	956	**915**
51	2341	2330	**1385**	2353	2330	2765	2338	2336
52	**1606**	1687	1898	2005	5632	4451	3551	3607
22	5006	1792	1807	1797	1795	1792	1797	**1750**
24	5022	4827	4817	4753	3887	3891	4807	**3877**
59	3772	3896	1085	1359	1104	1474	1105	**1103**

4 Conclusions

The obtained results confirm the assumption that emotional expressions are very complex and require a very detailed and accurate approach to analyses aimed at remote identification of emotions. Given the fact that, not all parts of the face react at the same time and at the same intensity (as has been proven), it is not correct to analyse emotions from individual photos. One image is not a material that can serve as data for a reliable diagnosis of emotion. To get information about the emotional state of a person on the basis of image analysis, it is necessary to observe before the change of emotion, then we have a complete database that allows accurate and thorough analysis of the expression and tracking of changes that take place over time.

It is important to use tools to track changes to get relevant information about the emotions in people with the accuracy of the order of thousandths of a second. In this case, we get the opportunity to catch micro-expressions lasting a few milliseconds that are not controlled by the person being tested. And it is these emotions that reflect the true feelings. The collected material gave another insight into the observations of changes occurring on the face. Analysis for a single movie of several thousand frames, lasting a few seconds, were very time consuming. Therefore, when planning accurate emotional states analysis using high-speed cameras, it is necessary to consider the material handling time. This is of great importance in planning further research related to the analysis of the research material obtained.

References

1. M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, K. Scherer, IEEE Trans. Syst. Man, Cybern. Part B **42(4)**, 966 (2012)

2. D. Datcu, L. J. M. Rothkrantz, *Emot. Recognit. A Pattern Anal. Approach,* chapter 16 (2014)

3. K. N. Ochsner, J. J. Gross, Trends Cogn. Sci., **9(5)**, 242 (2005)

4. T. Pfister, X. Li, G. Zhao, M. Pietik, *Proceedings of 2011 International Conference on Computer Vision,* 1449 (2011)

5. Russell, Tamara A., Chu, Elvina, M. L. Phillips, Br. J. Clin. Psychol. **45(4)**, 579 (2006)

6. Y. J. Liu, J. K. Zhang, W. J. Yan, S. J. Wang, G. Zhao, X. Fu, IEEE Trans. Affect. Comput. **7(4)**, 299 (2016)

7. D. Lundqvist, A. Flykt, A. Öhman, *The Karolinska directed emotional faces (KDEF)* (CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998)

8. X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1 (2013)

9. Wang, Y., J. See, R. C. W. Phan, Y. H. Oh, PLoS One **10 (5)**, e0124674 (2015)

10. K. Bobkowska, A. Janowski, M. Przyborski, *SGEM2015 Conference Proceedings* **2 (1)**, 1089 (2017)

11. M. Błażek, K. Bobkowska, M. Kaźmierczak, A. Janowski, M. Przyborski, J. Szulwic, *SGEM2017 Conference Proceedings* **17 (21)**, 655 (2017)

12. M. Blazek, J. Janowski, M. Kazmierczak, M. Przyborski, J. Szulwic, *ICERI2014 Proceedings*, 3787 (2014)

13. P. Viola, M. Jones, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **1**, I-511 (2001)