Identification of Emotions Based on Human Facial Expressions Using a Color-Space Approach

Zdzisław Kowalczuk* and Piotr Chudziak

Gdańsk University of Technology, Faculty of Electronics
Telecommunications and Informatics
Department of Robotics and Decision Systems
Gdańsk, Poland
kova@pg.gda.pl

Abstract. HCI technology improves human-computer interaction. Such communication can be carried out with the use of emotions that are visible on the human face since birth [1]. In this paper the Emotion system for detecting and recognizing facial expressions, developed in the MSc work [2], is presented. The system recognizes emotion from webcam video in real time. It is based on color segmentation and morphological operations. The system uses a cascade of boosted classifiers based on Haar-like features, to locate the face and to reduce the searched area for characteristic points. For identification purposes, the Emotion system uses an expanded action unit EAU, based on a facial action coding system, FACS [3, 4].

Keywords: Digital Image Processing, Segmentation, Emotion Recognition.

1 Introduction

Research in facial expressions was carried out already, in the XIX century, by Charles Darwin, the founder of the theory of evolution. He claimed [5] that facial expressions are universal and understood equally by people from different environments. Unfortunately, the theory did not meet with the approval of the majority of scholars.

It was not until a century later that Ekman's research [6] had shown that emotions are biologically conditioned and similar in every culture, and can be isolated into basic emotions: anger, fear, sadness, joy, disgust, surprise. In addition, he identified 43 facial muscles which allow humans to express these emotions. A facial action coding system (FACS) was created for coding particular facial movements.

According to the research of psychologist Albert Mehrabian on the essence of non-verbal communication [7, 8] there has been the '7-38-55' principle formulated [9], under which interpersonal communication is divided into three parts: non-verbal communication (gestures, posture, countenance/face/facial appearance) which takes 55% of participation in communication, the tone of voice – 38%, and speech content – solely 7% , respectively. Therefore, communication is mainly based on the non-verbal transmission and the tone of voice, ignoring the actual content of speech.

These studies have shown that the best, versatile and intelligent, way [10] of communication between man and a computer system consists of recognizing facial expressions.

2 Extended AU

FACS is a well-known system for coding changes in facial appearance. It defines action units (AU), with the use of which one can describe any facial expression that Man can take [11]. The system is closely related to facial anatomy, and makes the movements of facial muscles easy to observe [12].

The FACS system also describes the intensity of each AU by assigning to it a letter from A to E (Fig. 1), where A corresponds to traces of facial expression, whereas E represents a strong expression (this feature, though, will not be used here).

For identification purposes an extended list of action units EAUs (Extended Action Units) has been defined in Tab. 1, based on Cohn-Kanade's database [13, 20]. EAUs allow us to easily distinguish the characteristic features for both sides of face (left and right), as well as to combine two basic units AU as one synthetic unit.

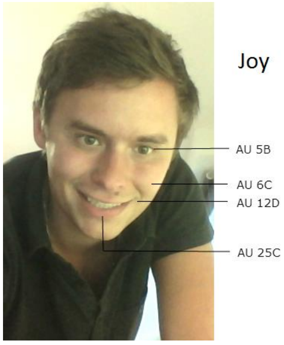


Fig. 1. Description of joy with AUs

Table 1. List of basic and extended action units (EAU) with description

AU EXTENSION	EAU LOGO	DESCRIPTION
AU1AU2R	**EAU1**	Right eyebrow raiser
AU1AU2L	**EAU2**	Left eyebrow raiser
AU4	**EAU4**	Eyebrows lowered and drawn together
AU5	**EAU5**	Upper lid raiser
AU9	**EAU9**	Nose wrinkler, nostrils extended
AU10L	**EAU10**	Upper left lip raiser, nasolabial furrows deepened
AU10R	**EAU11**	Upper right lip raiser, nasolabial furrows deepened
AU12	**EAU12**	Lip stretcher
AU13R	**EAU13**	Right lip corner raiser, lip tightener
AU13L	**EAU14**	Left lip corner raiser, lip tightener
AU15AU20	**EAU15**	Lip corner depressor, lip stretcher
AU23	**EAU23**	Lip tightener
AU26	**EAU26**	Jaw drop
AU44	**EAU44**	Slit

3 Decision Table

A simple decision table is used to assess the facial expression. Such a conversion table contains the rules that must be met by a given emotion. The Emotion system detects 9 emotional states as shown in Tab. 2. It contains the facial appearances represented by the combined EAUs. Each EAU is responsible for a specific unit, has a value 1 if it occurs on the face; 0 in the other case; or -1 for a facial expression in which it is possible to distinguish positive and negative movement directions.

Table 2. Decision table

EMOTION	DECISION
ANGER	EAU4 > 0 ∧ (EAU10 > 0 ∨ EAU11 > 0 ∨ EAU23 > 0)
DISGUST	EAU9 > 0
JOY	EAU14 > 0 ∧ EAU13 > 0
SADNESS	EAU15 > 0 ∧ EAU5 < 1
FEAR	EAU5 > 0 ∧ EAU4 > 0 ∧ (EAU15 > 0 ∨ EAU12 > 0 ∨ EAU26 > 0)
SURPRISE	(EAU5 > 0 ∧ EAU26 > 0) ∧ (EAU4 < 1 ∨ EAU15 < 1)
CONTEMPT	(EAU10 > 0 ∧ EAU11 < 1) ∨ (EAU10 < 1 ∧ EAU11 > 0) ∧ (EAU13 < 1 ∨ EAU14 < 1)
WONDERMENT	(EAU1 > 0 ∧ EAU2 < 1) ∨ (EAU1 < 1 ∧ EAU2 > 0) ∧ EAU4 < 1
NEUTRAL	Other

4 Face Detection

In order to speed up the detection and analysis of the emotion expressed by a spotted face; the system, after capturing a sample frame from a camera, performs image preprocessing and face location to limit the searched area of facial features (Viola-Jones' method with a Haar classifier). It is characterized by a high detection rate of positive image samples (containing a face), and a very low rate of false detection (empty samples, without a face) [14].

5 Image Processing

The Emotion system performs a few phases of image processing and analysis: preprocessing, segmentation, and identification along with feature analysis and classification. The first phase consists of noise suppression with a median filter (dramatically reducing impulse noise, with some blurring effects), conversion to grayscale, image scaling (to accelerate the algorithm and to reduce memory consumption), histogram equalization, and face location.

5.1 Segmentation

Upon the Viola-Jones face detection, in the second phase called segmentation, the image is empirically divided into the areas which represent the places of the occurrence of a facial feature. As a result, the searched range for specific points is limited, speeding up the analysis and detection. Each indicated region is subject to processing which allows for efficient binarization of the sought object.

Image transformation used in the Emotion are: normalization, Gaussian blur, Laplace transform, morphological operations and Otsu's thresholding method.

5.2 Feature Extraction from the Eye Area

The region of the potential occurrence of eyes is extracted, and subject to processing responsible for finding pupils. The extraction of pupils is based on [16], where the image is inverted and converted to a gray scale, and then the pupil's area is found as the highest brightness; which is easy to detect by thresholding (as illustrated in Fig. 2).

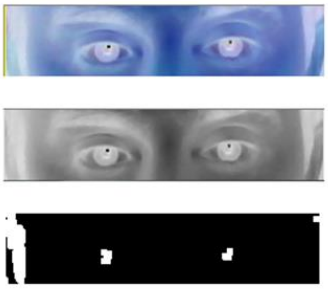


Fig. 2. Finding pupils: inverted, in greyscale, and binary images

To improve the effect of binarization and to facilitate the extraction of facial features, morphological dilation is used before, and after, thresholding. Information about the position of pupils is beneficial to the emotion analysis. One of the benefits is the information about the position of the mouth in the image, as described by Kubanek [17]. After extracting the pupils, the Euclidean distance (d) between their assessed means (centers) is calculated. Then the center s of this distance is determined, and a perpendicular at this center to the line between the pupils (length d) is created, where the end point has the coordinates (x_m, y_m):

$$x_m = x_s + (y_l - y_r)\ , \tag{1}$$

$$y_m = y_s + |x_l - x_r|\ , \tag{2}$$

where x_r and y_r are the right pupil coordinates, x_l and y_l are the left pupil coordinates, and $s(x_s, y_s)$ describes the center point between the pupils.

There are two benefits of the information about the location of pupils. The first is the fact that the pupillary distance does not change, but is the same even if you change the line of sight. Therefore, it can be used to calculate the angle of inclination of the head to a side. Any area including the specific feature is subject to rotation in accordance to the head tilt:

$$\alpha = \operatorname{arctg}\left(\frac{y_l - y_r}{x_r - x_l}\right), \tag{3}$$

wherein the resulting inclination angle α is in radians.

The second benefit is the knowledge of the probable areas where the upper and lower eyelids can be found. In these areas the R component of the image is subtracted from the B component. After applying Gaussian blur with a 5x5 mask, and morphological operations using a 3x3 structural element, thresholding operations, a median blur 3x3, image negation and morphological dilation operation 3x3; they all give the effect shown in Fig 3.1.

5.3 Features Extraction from the Mouth Area

In the designated mouth area, the extraction of all required characteristic points is carried out, which allows the identification of possible facial expressions performed by the mouth. These points are the position of the lower lip and corners of the mouth. The corners extraction proceeds according to the following steps: converting the color image into grayscale, Gaussian blur with a 7x7 mask (Fig. 3.2.b), adaptive thresholding, morphological opening with a 3x3 structural element, median blur 3x3, and image negation. An example of the results of these operations is shown in Fig 3.

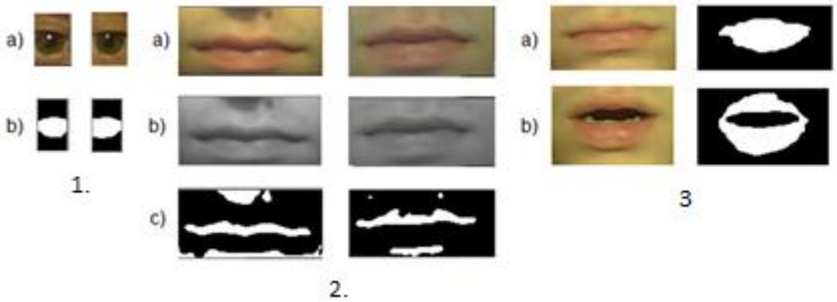


Fig. 3. Mouth extraction: (1) binary image of the left and right eyes; (2) detection of mouth corners; (3) lip binarization

The lip extraction is not a trivial task, especially in color. Often, the lack of contrast, and variable lighting, can degenerate the detection. In bad conditions, the difference between the mouth color and skin color is small. However, from a practical point of view, this method is quite simple and fast, and may be effective after increasing the contrast between the skin and the mouth.

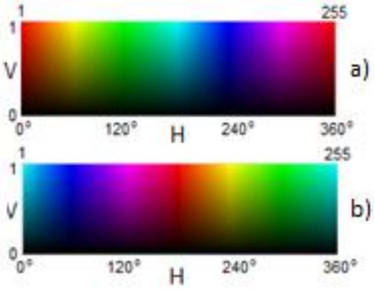


Fig. 4. Characteristics of color H and its cyclic shift

At the first stage of lower-lip detection, the image is transformed from the RGB space to the HSV space. This is a certain obstacle for the selection of thresholding values in the segmentation process, because the lips have a mainly red tint/color; which, as can be seen in Fig 4, may reach a value of 1 to 20 or 235 to 255 in digital representation. This variation of the red shades may result in bad segmentation. Hence the cyclic 180° shift is used for the Hue component. In the resultant new scale the Hue component represents the red color and its shades, in the range of 103 to 153.

Image segmentation uses a thresholding method with the following values:

$$\mathrm{Thr}_1(Ht_1, St_1, Vt_1) = (H\mu - H\sigma, S\mu - S\sigma, V\mu - V\sigma)\,, \tag{4}$$

$$\mathrm{Thr}_2(Ht_2, St_2, Vt_2) = (H\mu + H\sigma, S\mu + S\sigma, V\mu + V\sigma)\,, \tag{5}$$

where Thr_1 and Thr_2 describe the lower and the upper limits of the threshold, *H*, *S*, and *V*, are the principal components of the HSV space, μ is the average of the pixel values for each component, and σ is the standard deviation of all pixels contained in a given mask for each component.

5.4 Feature Extraction from the Nose Area

Features in the region of the nose provide information about expressed emotions, such as disgust or contempt. Therefore, the system does search for characteristic features, which correspond to the position of nostrils, and detects the appearance of nasolabial folds (Fig. 5). After defining the area of the nose in the process of segmentation, the following acts are performed: converting the image to grayscale, increasing the contrast of the image, raising the image resolution two times (bicubic interpolation), morphological erosion 3x3, median blur 5x5, thresholding, image negation and morphological dilation with a 3x3 structural element. The position of the nostrils, gained in this way, allows us to specify the exact areas of the occurrence of the nasolabial folds which are subjected to: conversion to grayscale, Gaussian blur with a 5x5 mask, Laplace transformation, median blur 3x3, and Otsu’s thresholding.

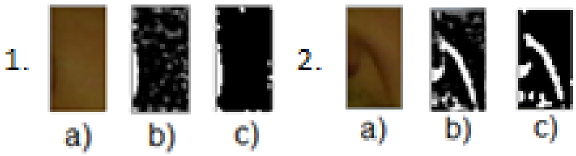


Fig. 5. Nose extraction in neutral (1) and contempt (2) emotions: (a) original image; (b) image after GL filter; (c) thresholded image.

For the detection of the nasolabial folds in the binary image, the number of non-zero pixels is calculated as

$$Pix_{count} = \sum_{I:scr(I)\neq 0} 1 \ , \tag{6}$$

where $scr(I)$ is one channel input image. This number is compared to a specified detection threshold Thr, above which we talk about the appearance of nasolabial folds (otherwise – about their absence).

There are also other characteristic points in the area of the nose, like the nasal alar, that provide additional information about emotions. To detect them, image normalization is completed, which is followed by: increasing the image resolution two times (with the use of bicubic interpolation), Gaussian blur 5x5, Laplace transformation, median blur 5x5, Otsu's thresholding, morphological opening and closing with 5x5 structural elements. The obtained effect is illustrated in Fig. 6.

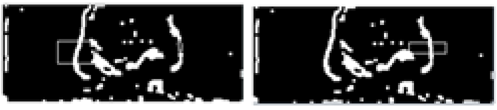


Fig. 6. Detection of nasal alar – regions prepared for analysis

5.5 Feature Extraction from the Eyebrow Area

Selected regions which may include eyebrows are cut and submitted to further processing. To extract eyebrows, processes taken from [18] are modified and implemented using: median blur 5x5, image conversion to grayscale, image negation, morphological opening 5x5, subtract image from negated image, image normalization, Otsu thresholding, morphological dilation 3x3 and median blur with a 3x3 mask. After passing through these steps one obtains a binary image ready for the analysis of features. The stages of eyebrow extraction are illustrated in Fig. 7.

Another characteristic feature that can be found in the area of the eyebrows, are the wrinkles between the eyebrows, called the wrinkles of the lion. These wrinkles (illustrated in Fig. 8) are the most characteristic features that allow you to recognize human emotions. Via their detection and analysis, the system can identify at least two emotions: fear and anger. After determining the position and dimensions of the isolated area, it is treated with a GL filter, according to the following scheme: image conversion to grayscale, Gaussian blur with a 7x7 mask and the Laplace transform.

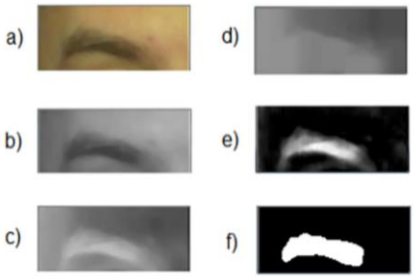


Fig. 7. Eyebrows extraction: (a) initial blur; (b) grayscale image; (c) negated image; (d) image after morphological opening; (e) image resulting from subtraction (d – c); (f) binary image after thresholding, dilation and blur.

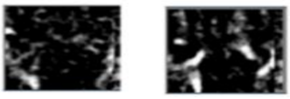


Fig. 8. Binary image between eyebrows: without (left) and with (right) wrinkles

6 Identification and Analysis

The system determines here the position of characteristic points. Their detection is accomplished by locating the sample in the image, measuring its geometrical properties (dimensions), and calculating its geometric moment (mass center). The system is capable of locating 20 characteristic facial features and 3 areas of the appearance of wrinkles, which are important in terms of emotion-recognition.

7 Classification

The last phase of image processing is classification. The FACS system is applied in the detection of the expression of emotion. To discover the occurrence of an action unit, information about the position of the characteristic facial points (Fig. 9) is utilized. By determining the distances between these points, and by calculating the sum of non-zero pixels in the wrinkle area, we identify facial appearances. To reduce the impact of changes in the distance of the face from the computer screen on the analyzed parameters, normalization is implemented using the distance between the pupils of the eyes as a referencing value, which is the greatest distance of all calculated parameters d_n $(n = 1, \dots, 10)$. Therefore, each normalized parameter is a fraction of the distance between the two pupils. It is also obvious that the distance d_{10} is invariant and does not change with the movement of the eye.

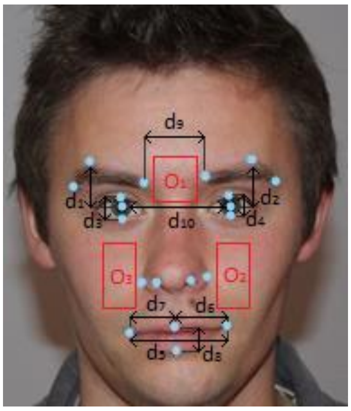


Fig. 9. Distance between characteristic points/features.

For each characteristic feature an emblem is ascribed, whose value depends on the expression performed with the use of this particular feature. Each emblem takes its values from the set {-1, 0, 1}, according to the thresholding detection. These values indicate the appearance of the different features concerning the movable face features; such as eyebrows or lips, the lion's wrinkles or the nasolabial furrow, and their movement direction (stretch, tighten, lift, pull down).

Table 3. Detection of the extended action units using emblems

Extended AU	Condition	Description
EAU1	$\mathbf{E}_2(\mathrm{d}_2) > 0$	Right eyebrow raised
EAU2	$\mathbf{E}_1(\mathrm{d}_1) > 0$	Left eyebrow raised
EAU4	$\mathbf{E}_3(\mathrm{O}_1) > 0$ $\mathbf{E}_2(\mathrm{d}_2) < 1 \wedge \mathbf{E}_1(\mathrm{d}_1) < 1$	Eyebrows lowered and drawn together
EAU5	$\mathbf{E}_4(\mathrm{d}_3) > 0 \vee \mathbf{E}_5(\mathrm{d}_4) > 0$	Eyes wide opened
EAU9	$\mathbf{E}_{11}(\mathrm{O}_2) > 0 \wedge \mathbf{E}_{12}(\mathrm{O}_3) > 0$	Nose wrinkled, nostrils stretched
EAU10	$\mathbf{E}_{11}(\mathrm{O}_2) > 0$	Upper left lip lifted, nasolabial furrow deepened
EAU11	$\mathbf{E}_{12}(\mathrm{O}_3) > 0$	Upper right lip lifted, nasolabial furrow deepened
EAU12	$\mathbf{E}_7(\mathrm{d}_5) > 0$	Mouth corners pulled to the sides
EAU13	$\mathbf{E}_9(\mathrm{d}_6) > 0$	Right mouth corner raised, lips pursed
EAU14	$\mathbf{E}_8(\mathrm{d}_7) > 0$	Left mouth corner raised, lips pursed
EAU15	$\mathbf{E}_8(\mathrm{d}_7) < 0 \vee \mathbf{E}_9(\mathrm{d}_6) < 0$	Mouth corners pulled down, lips stretched
EAU23	$\mathbf{E}_7(\mathrm{d}_5) < 0$	Lips pressed and crest together (pouted)
EAU26	$\mathbf{E}_{10}(\mathrm{d}_8) > 0$	Jowl dropped
EAU44	$\mathbf{E}_4(\mathrm{d}_3) < 0 \wedge \mathbf{E}_5(\mathrm{d}_4) < 0$	Eyes narrowed

Using such utilized emblems, the Emotion system proceeds to identify the extended action units (EAUs) listed in Table 3.

8 Tests

The Emotion system has been tested in two lighting conditions. We paid special attention to the system's behavior in its normal exploitation, i.e. without the use of light sources directly aimed at the human face, and without manipulating the surroundings of the human. All research results were obtained in real-time experiments conducted by the author of the MSc thesis and based on the facial expressions of this student [2].

After initialization and calibration of the system, in a test verifying the continuity of tracking of the characteristic points, the human subject began to make moves of his head from side to side, and next forward and backward. It turned out that the movements to the side have a greater impact on the detection quality than those receding and impending to the computer camera. One of the reasons for this behavior is the property of the web camera used to capture the image, which automatically adjusts to the light conditions through software brightening or dimming of the image.

The Emotion system had, however, a problem with the detection of characteristic points in moments when the lighting deteriorated the image significantly. It could not correctly locate the face elements. Recalibration of the system, in such situations, easily eliminated this problem.

Testing the accuracy of the system in the theoretically better conditions: that is, in natural lighting, leads; in general, to similar results as under artificial lighting. In this case, however, the characteristic points were lost more frequently. The reasons for this were foreign objects standing out in the search area. In particular, the elements located near the right eyebrow, such as wrinkles on the forehead and the hair on the temple, which interfere with proper detection. As mentioned earlier, identification of the expressed emotion is based on a proper decision system. Thus, the accuracy of interpretation of individual emotions depends on precise definitions of the facial appearance. Certainly, wrongly assigned values to the emblems imply that the system will misinterpret emotions.

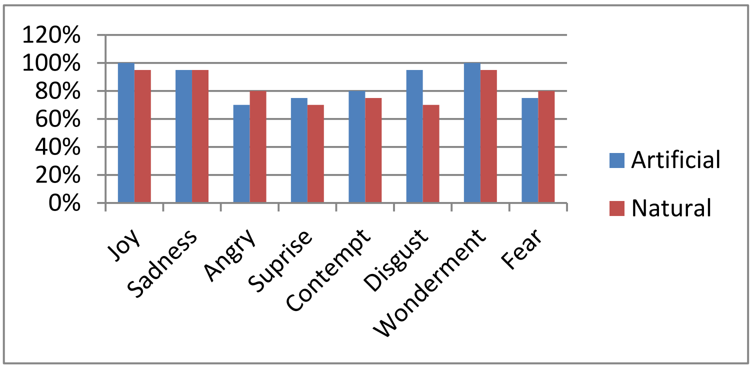


Fig. 10. Recognition quality of particular emotions in different lighting condition

The decision rules were chosen based on the studies available from [19] and the results of the research conducted on countenances and emotions in [2].

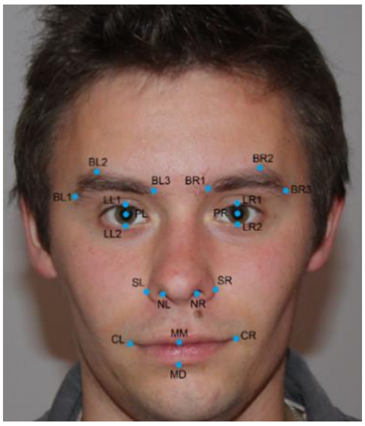
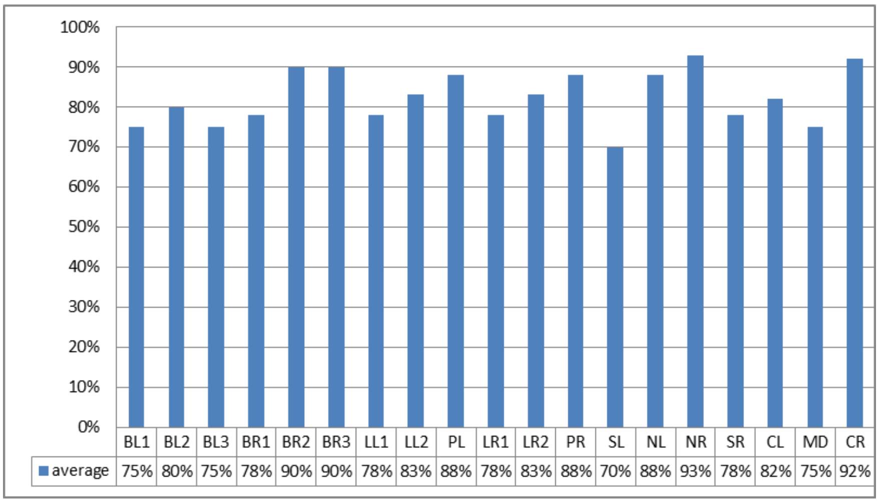


Fig. 11. Face identifiers

On the basis of the collected data, it is possible to determine the accuracy of the system depending on the prevailing lighting. On average, under natural lighting conditions, the detection accuracy of certain emotions is 4% lower than in the case of artificial lighting (Fig. 10).

Additional tests were performed to authenticate the tracking quality of the facial characteristic points. The tests consisted of performing specific facial movements, each being a characteristic gesture for a concrete feature. For this purpose auxiliary identifiers for each characteristic point, shown in Fig. 11, are defined.



	BL1	BL2	BL3	BR1	BR2	BR3	LL1	LL2	PL	LR1	LR2	PR	SL	NL	NR	SR	CL	MD	CR
■ average	75%	80%	75%	78%	90%	90%	78%	83%	88%	78%	83%	88%	70%	88%	93%	78%	82%	75%	92%

Fig. 12. Average tracking quality of the facial feature points

From the obtained results shown in Fig. 12, it can be concluded that the system detects effectively the characteristic points located on the right side of the face. The effect concerning the left side can be explained by the lighting being not sufficiently uniform. Moreover, it has been mentioned that the characteristic points are often lost for facial gestures supported by movements like lowering, dropping and squinting. This is because, for some facial expressions, the objects analyzed in a binary image can appear to be too close to each other, thereby causing interference, which results in a wrong interpretation of the position of the characteristic points. This phenomenon occurs, for example, with pixels representing the left eyebrow: which, after lowering, can be connected with the pixels responsible for the nose wrinkles or upper eyelid.

9 Conclusions

While working on Emotion, a range of methods and tools for the detection of facial expressions have been applied and tested. Unfortunately, most of them had to be rejected, leaving only a few methods that met the most desirable criteria for speed, and efficiency, of extracting information. These allow capable tracking of facial gestures based on the movement of pupils, eyelids, eyebrows and lips (which are also applied in most computer facial animations).

Research on the effectiveness of detection and identification of emotions shows that the Emotion system works best in areas with limited natural light. Uniform illumination on the face, without unwanted shadows and contrast areas, allows the program to work flawlessly. In the external ambient conditions, it behaves a bit differently. Equal lighting on all surfaces (including background and foreground) contributes to the small differences in contrast.

Analysis of the obtained results leads to the conclusion that the Emotion system can effectively detect emotions accompanied by very strong expression (of those emotions). When decreasing the strength of the expression, the system tends to make more mistakes in identifying emotion. The most correctly identified emotions are those which are not represented by a complex execution of facial expressions, such as Joy (97.5%), Sadness (95%) and Surprise (97.5%). On the contrary, the system has difficulties with identification of Anger (75%) and Surprise (72.5%).

Experience shows that the Emotion system can be successfully used in various fields of human activity, although certainly it still requires some improvements. We intend, for instance, to move away from the color-space approach, as it is a quite challenging task for digital image processing; because even a small change in lighting or visible derogations in the environment can lead to identification errors. With the calibration functionality built into the Emotion system, such errors can be avoided, although this is not the most convenient solution for automatic recognition systems for Human emotion.

References

1. Lyons, M.J., Bartneck, C.: HCI and the face. HCI (2006)
2. Chudziak, P.: Analiza ekspresji mimiki ludzkiej twarzy za pomocą komputera. Master thesis (under the supervision of Prof. Z. Kowalczuk), Politechnika Gdańska (2015)
3. Ekman, P., Friesen, W.V.: Manual for the Facial Action Coding System. Consulting Psychologists Press (1977)
4. Ekman, P., Friesen, W.V., Hager, J.C.: Facial Action Coding System Investigator's Guide. A Human Face. Salt Lake City (2002)
5. Darwin, C.: The Descent Of Man, and Selection in Relation to Sex. John Murray (1871)
6. Ekman, P.: Facial expressions. Handbook of Cognition and Emotion. John Wiley & Sons Ltd. (1999)
7. Mehrabian, A., Wiener, M.: Decoding of inconsistent communications. Journal of Personality and Social Psychology, **6**(1), 109-114 (1967)
8. Mehrabian, A., Ferris, S.R.: Inference attitudes from nonverbal communication in two channels. Journal of Consulting Psychology, **31**(3), 248-252 (1967)
9. Mehrabian, A.: Silent Messages: Implicit Communication of Emotions and Attitudes. Wadsworth, Belmont, CA (1981)
10. Viola, P., Jones, M.: Rapid object detection using boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511-518 (2001)
11. http://niewerbalnie.info/ekspresja_twarzy.html (2016)
12. Przybyło, J.: Automatyczne rozpoznawanie elementów mimiki w obrazie twarzy i analiza ich przydatności do sterowania. Akademia Górniczo-Hutniczna im. St. Staszica, Kraków (2008)
13. Kanade, T., Cohn, J. F.: Cohn-Kanade AU-Coded Facial Expression Database. Robotics Institute, Carnegie Mellon University (2000)
14. Wilson, P.I., Fernandez, J.: Facial Feature Detection using Haar Classifiers. Consortium for Computing Sciences in Colleges (2006)
15. Wilson, P., Fernandez, J.: Establishing a face recognition research environment using open source software. ASEE Gulf-Southwest Annual Conference (2005)
16. Gupta, K.D.: Pupil or eyeball detection and extraction from eye image using C#. Code Project (2010)
17. Kubanek, M.: Metoda rozpoznawania audio-wideo mowy polskiej w oparciu o ukryte modele markowa. PhD thesis. Politechnika Częstochowska (2005)
18. Sohail, A.S.M., Bhattacharya, P.: Detection of facial feature points using anthropometric face model. Signal Processing for Image Enhancement and Multimedia Processing. vol. 31, ss. 189-200 (2006)
19. Friesen, W.; Ekman, P.: EMFACS-7: Emotional Facial Action Coding System. Unpublished manual. University of California, California (1983)
20. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. Proc. of FGR00 (2000)