Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 1175–1184

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Emotion recognition using facial expressions

Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak

Warsaw University of Technology, Warsaw, Poland.

pawel.tarnowski@ee.pw.edu.pl, marcin.kolodziej@ee.pw.edu.pl, andrzej.majkowski@ee.pw.edu.pl, remigiusz.rak@ee.pw.edu.pl

Abstract

In the article there are presented the results of recognition of seven emotional states (neutral, joy, sadness, surprise, anger, fear, disgust) based on facial expressions. Coefficients describing elements of facial expressions, registered for six subjects, were used as features. The features have been calculated for three-dimensional face model. The classification of features were performed using k-NN classifier and MLP neural network.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: facial expression, emotion recognition, action units, computer vision, k-NN, MLP

1 Introduction

Facial expressions play an important role in recognition of emotions and are used in the process of non-verbal communication, as well as to identify people. They are very important in daily emotional communication, just next to the tone of voice [1]. They are also an indicator of feelings, allowing a man to express an emotional state [2,3]. People, can immediately recognize an emotional state of a person. As a consequence, information on the facial expressions are often used in automatic systems of emotion recognition [4]. The aim of the research, presented in this article, is to recognize seven basic emotional states: neutral, joy, surprise, anger, sadness, fear and disgust based on facial expressions.

Man's face, as the most exposed part of the body, allows the use of computer vision systems (usually cameras) to analyze the image of the face for recognizing emotions. Light conditions and changes of head position are the main factors that affect the quality of emotion recognition systems using cameras [5]. Especially sensitive for these factors are methods based on 2D image analysis. Methods in which 3D face models are implemented are far more promising.

In our experiments we used Microsoft Kinect for 3D face modeling mainly because of its low price and simplicity of operation. Kinect has small scanning resolution, but a relatively high rate of image registering (30 frames/s). It has an infrared emitter and two cameras. One of the cameras record visible light, while the other operates in infrared and is used for measuring the depth [6]. Infrared rays reflected from the user's body allow creating a 3D model of a face. The model (Candid3 [7]) is based

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.025

on 121 specific points of the face, recorded by the Kinect device. These points are arranged on characteristic positions on the face such as the corners of the mouth, nose, cheekbones, eyebrows, etc. A set of characteristic points of the face registered in 2D space is shown in Fig. 1.

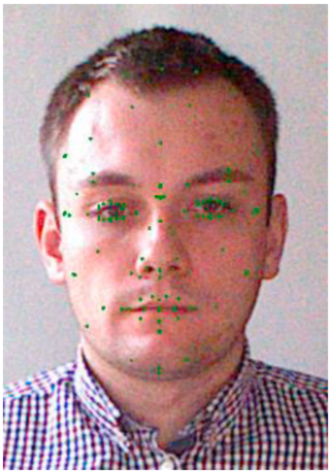


Figure 1: Characteristic points on the face

Spatial coordinates of the points are stored in a form of a matrix. Coordinate system (x, y, z), as defined in Kinect device, is shown in Fig. 2. Changes in facial expressions resulting from the activity of specific muscles [8] have been defined in the developed by Ekman and Friesen FACS system (Facial Action Coding System) [9] in the form of special coefficients - Action Units (AU). For example, the movement of the inner part of eyebrow, for which frontal cranial vault muscle is responsible, is described by the coefficient "Action Unit 1" called *Inner Brow Raiser*.

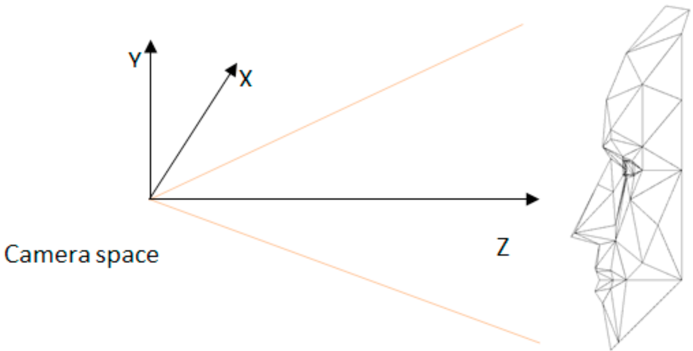


Figure 2: Kinect coordinate system **[10]**

Kinect device provides six Action Units (AU) derived from the FACS system. The Action Units may be used to describe emotions either separately or in combinations. AU take values between –1 and +1, and carry information about: AU0 - upper lip raising, AU1- jaw lowering, AU2 - lip stretching, AU3 - lowering eyebrows, AU4 - lip corner depressing, AU5 - outer brow raising.

2 Materials

Six men aged 26-50 took part in the experiment. Each experiment participant took a seated position, at a distance of 2 meters from the Kinect unit. A participant task was to play mimic effects according to instructions on a computer screen. The instructions contain name of the emotional state and a picture of an actor performing the corresponding mimic effect (we used images from KDEF database [11]). The aim of presentation of pictures was to make easier of playing a proper mimic effect. Instruction was displayed for 5 seconds. The participants were asked to play a mimic effect for the duration of the instruction. Between emotional states there were 3-second breaks. In Fig. 3 there are presented sample images of actors whose facial expressions corresponded to emotional states such as neutral, joy, surprise, anger, sadness, fear, disgust.

Figure 3: Facial expressions presented to users **[11]**

Each subject participated in two sessions. Each session consisted of three trials in which a participant mimicked, in sequence, all seven examined emotional states. As a result, forty two (2×3×7=42) 5-second sessions were registered for each user. The entire database contained a total of 252 (6×42) facial expressions. The hierarchical structure of the experiment is shown in Fig. 4.

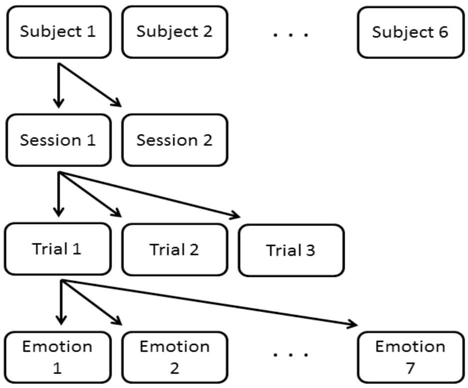


Figure 4: The structure of the experiment

3 Methods

In the classification process, six action units (AU), calculated by the Kinect device, were used as features. Table 1 shows exemplary values of AU for facial expressions of one of the participant. The images correspond to emotional states (ES): neutral, joy, surprise, anger, sadness, fear, disgust.

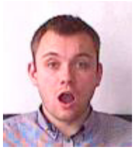
ES	neutral	joy	surprise	anger	sadness	fear	disgust
							
AU0	0.21	0.77	-0.10	0.30	0.17	-0.11	0.91
AU1	-0.06	0.09	0.60	-0.07	-0.04	0.20	0.13
AU2	-0.25	1.00	-0.49	0.06	-0.37	-0.60	0.88
AU3	-0.21	0.00	-0.13	0.04	-0.09	-0.17	0.00
AU4	-0.04	-0.47	0.58	-0.19	-0.02	0.28	-0.32
AU5	-0.23	-0.30	0.10	-0.34	-0.27	-0.02	-0.39

Table 1: The facial expressions and corresponding AU

An exemplary distributions of AU2 and AU5 obtained for subject #3 are shown in Fig. 5. Even preliminary analysis of the distribution of the features allows us to see the ability to distinguish some emotions. We decided to test the possibility of automatic recognition of emotions using AU. The experiments were conducted using k-NN and MLP classifiers [11].

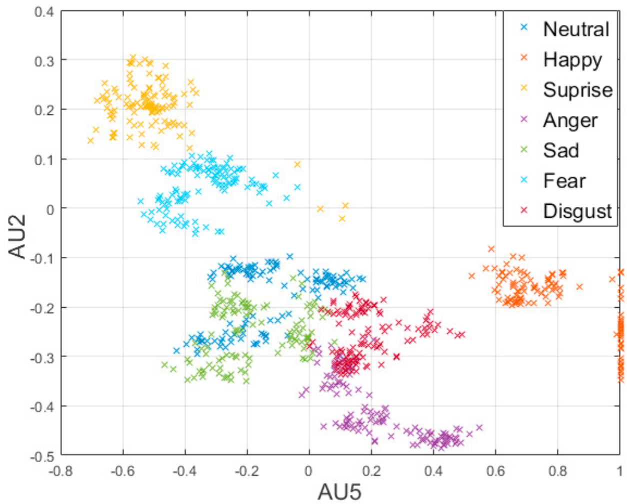


Figure 5: AU2 and AU5 distribution

The user response time and the time to prepare a proper facial expression could have a negative impact on the quality of the features in the first second of signal registration. Therefore, only AU recorded during the last four seconds of presentation of the emotional state were used to teach a classifier. We used nearest neighbor classifier (*3*-NN) and two-layer neural network classifier (MLP) [12] with 7 neurons in the hidden layer. The structure of the used neural network is shown in Fig. 6. The input of the network were six AU. The output was one of the seven emotional states.

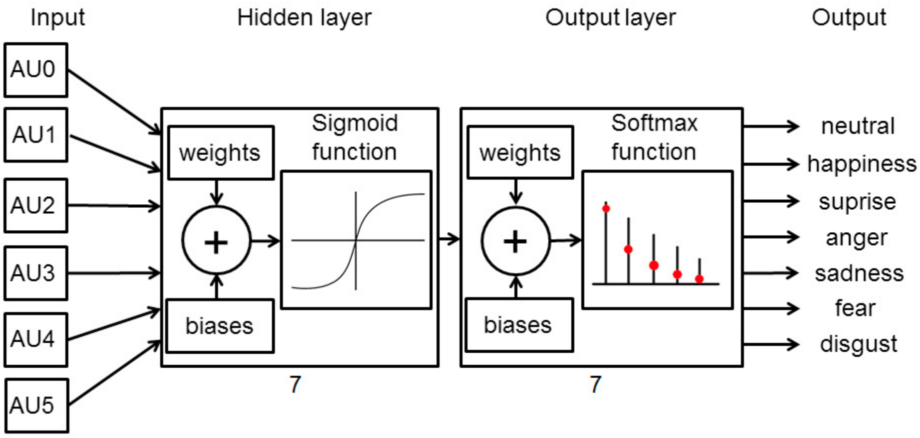


Figure 6: The neural network structure

4 Results

We tested two ways to recognize emotions: a) subject-dependent - for each user separately and b) subject-independent - for all users together. In both cases, for 3-NN classifier, data were randomly divided on the teaching part (70%) and the testing part (30%) and for MLP into three groups: teaching (70%), testing (15%) and validation (15%). Neural network was trained using back propagation algorithm with conjugate gradient method [13]. The results of the classification for subject-dependent case are shown in Table 2.

Subject	MLP	3-NN
1	0.94	0.97
2	0.96	0.96
3	0.90	0.98
4	0.74	0.90
5	0.96	0.96
6	0.93	0.97
Average	**0.90**	**0.96**

Table 2: The results of the subject-dependent classification

Recognition of emotions based on facial expressions for all users (subject-independent) is much more useful and versatile than for an individual user (subject-dependent). In subject-independent approach, the classifier accuracies (CA) for 3-NN and MLP algorithms were respectively 95.5% and 75.9%. The results are very good, especially for 3-NN classifier. For that case, in order to determine which emotions are the easiest and which the most difficult to distinguish, it is necessary to calculate the confusion matrices. Confusion matrices for 3-NN and MLP classifiers are shown in Tables 3 and 4.

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	425	2	1	3	10	0	1
joy	9	421	0	2	6	0	5
surprise	1	1	429	0	0	11	0
anger	7	1	0	428	1	0	6
sadness	20	4	0	2	416	1	0
fear	5	0	19	0	6	412	1
disgust	2	2	1	9	2	0	427

Table 3: Confusion matrix for 3-NN classifier (CA=0.95)

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	1130	65	1	40	505	4	45
joy	61	1102	0	68	149	0	124
surprise	0	0	1056	4	0	194	13
anger	47	157	0	1317	30	0	90
sadness	193	41	15	4	726	77	21
fear	4	2	404	0	55	1201	3
disgust	41	109	0	43	11	0	1180

Table 4: Confusion matrix for MLP classifier (CA=0.75)

In the next stage of the study, we perform classification for a different division of the data (for learning and testing). Classification conditions, closer to the real ones, have been created through the use of a special, in some sense "natural", division of data. For subject-dependent classification, the data were divided into 6 subsets (2 sessions x 3 trials) with all 7 facial expressions (see Fig. 4). Five subsets were used for classifier teaching and one for testing. The average classification results for the MLP and 3-NN classifiers are shown in Table 5.

Subject	MLP	3-NN
1	0.75	0.70
2	0.80	0.74
3	0.71	0.69
4	0.57	0.48
5	0.79	0.76
6	0.74	0.85
Average	**0.73**	**0.70**

Table 5: The accuracy of the subject-dependent classification for "natural" division of data

For the case of subject-independent classification, data were divided into 12 subsets. Each subset included data coming from a single session of one user. Eleven subsets were used for classifier teaching and one (coming from single session) for testing. The averaged classification results are shown in Table 6.

No	Subject-Session	MLP	3-NN
1	1-A	0.74	0.67
2	1-B	0.76	0.57
3	2-A	0.76	0.68
4	2-B	0.85	0.70
5	3-A	0.65	0.64
6	3-B	0.76	0.63
7	4-A	0.60	0.36
8	4-B	0.55	0.31
9	5-A	0.80	0.77
10	5-B	0.78	0.72
11	6-A	0.81	0.80
12	6-B	0.67	0.68
	Average	**0.73**	**0.63**

Table 6: The accuracy of subject-independent classification for "natural" division of data

For user-independent classification the highest classification accuracy (73%) was achieved for MLP neural network. Cumulative confusion matrices, for that case, were created by summing up the results of all subjects. The matrices for 3-NN and MLP classifiers are shown in Tables 7 and 8.

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	881	125	2	155	340	48	101
joy	106	922	7	238	115	1	154
surprise	13	1	1135	8	8	390	13
anger	130	151	6	862	81	2	120
sadness	229	130	33	101	823	104	88
fear	41	0	220	5	88	871	4
disgust	76	147	73	107	21	60	996

Table 7: Confusion matrix for 3-NN classifier (CA=0.63)

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	1160	75	9	29	644	61	41
joy	81	1178	0	57	141	0	129
surprise	3	0	1153	4	2	426	10
anger	58	137	0	1346	44	0	88
sadness	122	13	5	0	561	75	2
fear	8	1	308	1	74	910	2
disgust	44	72	1	39	10	4	1204

Table 8: Confusion matrix for MLP classifier (CA=0.73)

5 Discussion

The use of AU describing facial expression, together with 3D modeling allows to obtain good results of classification. For all users, we obtained classification accuracy of emotions of 96% (3-NN), 90% (MLP) for random division of data. For "natural" division of data the classification accuracy was 73% (for MLP classifier). In the same case, for the 3-NN classifier we obtained a classification accuracy of 10% worse. This shows that neural networks have a good ability to generalize.

Presentation of the results in the form of confusion matrices enabled us to determine the emotions that were recognized with the lowest accuracy. The most difficult to recognize were: sadness and fear. They were often confused respectively with neutral and surprise emotions. This is probably caused by using only six AU.

To analyze the possibility of distinguishing between particular emotions we performed additional tests - classification in pairs. The results obtained for *3*-NN classifier are shown in Table 9. These results confirm that most mistakes occur between pairs: sadness-neutral and surprise-fear. Facial expressions for surprise and fear are very similar and are characterized by opened mouth and a raised eyebrow. Similar changes of the same AU affect much the classification accuracy. In the case of sadness and neutral emotions the deterioration of accuracy can be caused by too little change of AU4 coefficient that should best distinguish neutral and sad face expression.

The results obtained for subject # 4 are clearly worse than for the others. This subject wore glasses during signal registration. In this case, the Kinect was not able to properly record the AU3 and AU5 coefficients - describing lowering and raising eyebrows. Also facial hair or skin color of a user could affect the quality of emotion classification.

Emotions	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	-	0.86	0.94	0.85	**0.71**	0.93	0.85
joy	0.86	-	0.98	0.84	0.84	0.98	0.83
surprise	0.94	0.98	-	0.98	0.94	**0.75**	0.95
anger	0.85	0.84	0.98	-	0.87	0.97	0.84
sadness	**0.71**	0.84	0.94	0.87	-	0,89	0.86
fear	0.93	0.98	**0.75**	0.97	0.89	-	0.94
disgust	0.85	0.83	0.95	0.84	0.86	0.94	-

Table 9: The classification accuracy of emotions in pairs

The experiments were performed under strictly defined conditions and proper user position in relation to the Kinect unit. We have examined the impact of user movements and head rotations for proper AU registration. Kinect allows recording data for user's head orientation in relation to the device in the range from –180 to +180 degrees. In Fig. 7 there are shown the possible changes of head orientation in relation to the x axis: up and down movement (a), y axis: right-left rotation (b) and z axis: right-left tilt (c).

Changing the head orientation in relation to x and y axis may cause that a part of the user's face is not visible to the Kinect device. The authors examined the effect of changes of head orientation in relation to the axis x and y on AU values. For this purpose, the coefficients were recorded during head movements. The range of changes in relation to x-axis was from –10 to +5 degrees and to y-axis: –30 to +30 degrees. The changes of AU0 values within the specified range are shown in Fig. 8.

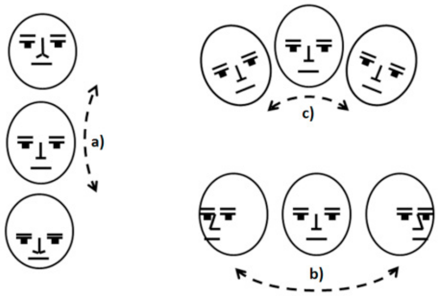


Figure 7: Rotation of the head relative to the camera

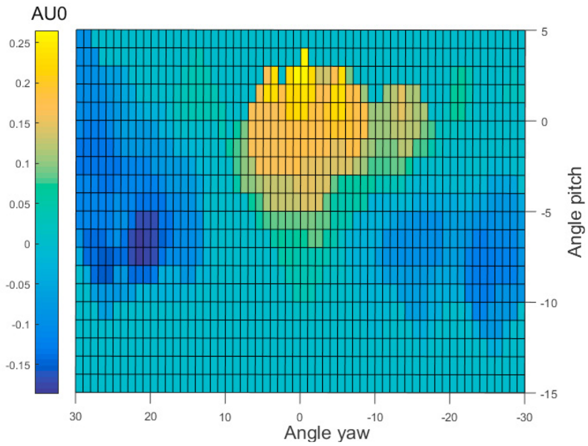


Figure 8: Changes of AU0 when changing the orientation of the head relative to the Kinect device

Changing the head orientation could significantly affect the value of AU coefficients. The impact of these changes on the classification process was tested for MLP. For this purpose extra session was registered, during which the examined person was not sitting in front of the device, but at a certain angle. In this case, the accuracy of classification for the MLP network was 54% and was about 20% worse in comparison with the previous results. Despite the use of 3D modeling, we observed a strong influence of user's position in relation to the Kinect unit for classification results.

6 Conclusion

In the carried out experiments, for 7 emotional states, we achieved a very good classification accuracy of emotions - 96% for random division of data and satisfactory classification accuracy - 73%, for "natural" division of data. This result was obtained for MLP classifier and "natural" division of data for all users (subject-independent). Experiments were carried out under the same conditions and at a fixed position of a user in relation to the Kinect unit. Certainly, the classification accuracy was influenced by the way users play specific facial expressions. In real conditions the classification accuracy can be affected by many additional factors. When you feel real emotions, facial expressions can vary greatly - may be exposed to a greater or lesser extent.

References

[1] Przybyło J., *Automatyczne rozpoznawanie elementów mimiki twarzy w obrazie i analiza ich przydatności do sterowania*, rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków, 2008.

[2] Ratliff M. S., Patterson E., Emotion recognition using facial expressions with active appearance models, *Proceedings of the Third IASTED International Conference on Human Computer Interaction*, ACTA Press, Anaheim, CA, USA, 2008, 138–143.

[3] Tian Y. I., Kanade T., Cohn J. F., Recognizing action units for facial expression analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (2001), no. 2, 97–115.

[4] Mao Q., Pan X., Zhan Y., Shen X., Using Kinect for real-time emotion recognition via facial expressions, *Frontiers Inf Technol Electronic Eng*, 16 (2015), no. 4, 272–282.

[5] Li B. Y. L., Mian A. S., Liu W., Krishna A., Using Kinect for face recognition under varying poses, expressions, illumination and disguise, *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, 2013, 186–192.

[6] Microsoft Kinect, https://msdn.microsoft.com/en-us/library/jj131033.aspx.

[7] Ahlberg J., *CANDIDE-3 - An Updated Parameterised Face*, 2001.

[8] Koelstra S., Patras I., Fusion of facial expressions and EEG for implicit affective tagging, *Image and Vision Computing*, 31 (2013), no. 2, 164–174.

[9] Ekman P., Friesen W., *Facial Action Coding System*, Consulting Psychologists Press, Stanford University, Palo Alto, 1977.

[10] Microsoft Kinect, https://msdn.microsoft.com/en-us/library/jj130970.aspx.

[11] Lundqvist D., Flykt A., Öhman A., The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, no. 1998,.

[12] Ogiela M. R., Tadeusiewicz R., Pattern recognition, clustering and classification applied to selected medical images, *Studies in Computational Intelligence*, 84 (2008), no. 117–151.

[13] Osowski S., *Sieci neuronowe do przetwarzania informacji*, 3rd ed., OWPW, Warszawa, 2013.