Metadata of the chapter that will be visualized in SpringerLink

Book Title	Image Analysis and Recognition	
Series Title		
Chapter Title	Facial Expression Based Emotion Recognition Using Neural Networks	
Copyright Year	2018	
Copyright HolderName	Springer International Publishing AG, part of Springer Nature	
Author	Family Name	**Yağış**
	Particle	
	Given Name	**Ekin**
	Prefix	
	Suffix	
	Role	
	Division	Faculty of Engineering and Natural Sciences
	Organization	Sabanci University
	Address	34956, Tuzla, Istanbul, Turkey
	Email	
Corresponding Author	Family Name	**Unel**
	Particle	
	Given Name	**Mustafa**
	Prefix	
	Suffix	
	Role	
	Division	Faculty of Engineering and Natural Sciences
	Organization	Sabanci University
	Address	34956, Tuzla, Istanbul, Turkey
	Email	munel@sabanciuniv.edu

Abstract	Facial emotion recognition has been extensively studied over the last decade due to its various applications in the fields such as human-computer interaction and data analytics. In this paper, we develop a facial emotion recognition approach to classify seven emotional states (joy, sadness, surprise, anger, fear, disgust and neutral). Seventeen action units tracked by Kinect v2 sensor have been used as features. Classification of emotions was performed by artificial neural networks (ANNs). Six subjects took part in the experiment. We have achieved average accuracy of 95.8% for the case in which we tested our approach with the same volunteers took part in our data generation process. We also evaluated the performance of the network with additional volunteers who were not part of the training data and achieved 67.03% classification accuracy.

Keywords (separated by '-')	Facial expression - Emotion recognition - Action Units (AU) - Kinect - Artificial Neural Networks (ANN)

Facial Expression Based Emotion Recognition Using Neural Networks

Ekin Yağış and Mustafa Unel(✉)

Faculty of Engineering and Natural Sciences, Sabanci University,
34956 Tuzla, Istanbul, Turkey
munel@sabanciuniv.edu

AQ1

Abstract. Facial emotion recognition has been extensively studied over the last decade due to its various applications in the fields such as human-computer interaction and data analytics. In this paper, we develop a facial emotion recognition approach to classify seven emotional states (joy, sadness, surprise, anger, fear, disgust and neutral). Seventeen action units tracked by Kinect v2 sensor have been used as features. Classification of emotions was performed by artificial neural networks (ANNs). Six subjects took part in the experiment. We have achieved average accuracy of 95.8% for the case in which we tested our approach with the same volunteers took part in our data generation process. We also evaluated the performance of the network with additional volunteers who were not part of the training data and achieved 67.03% classification accuracy.

Keywords: Facial expression · Emotion recognition
Action Units (AU) · Kinect · Artificial Neural Networks (ANN)

1 Introduction

Facial expressions are the integral part of human communication. Along with intonation and accentuation, facial expressions complement the communication by shaping the intended meaning. Facial expressions are the most basic and primitive reflections of human emotions, which are at the same time the most effective and inevitable part of any communication.

In human to human interaction, non-verbal part of the communication such as tone and facial expressions are analyzed internally. These supporters allow a human to comprehend the emotional state of a person and intended meaning of a communication. In 1968, Mehrabian pointed out that 55% of message conveying information about feelings and attitudes is transmitted through facial expressions [1]. Thus, facial expression recognition has been widely studied for measuring the emotional state of human beings.

Being the pioneers of ERFE field (Emotion Recognition via Facial Expressions), Ekman and Friesen proposed a discrete categorization for basic facial emotions under six groups: anger, disgust, fear, happiness, sadness and surprise; and the seventh subset can be identified as neutral state of expression [2]. In the

A. Campilho et al. (Eds.): ICIAR 2018, LNCS 10882, pp. 1–8, 2018.
https://doi.org/10.1007/978-3-319-93000-8_24

research, they have developed the FACS system (Facial Action Coding System) in which changes in facial expressions deriving from specific muscle activity have been described as special coefficients - Action Units (AUs). For instance, Orbicularis oculi and pars orbitalis muscles are active in the movement of cheeks. Upper movement of checks is defined by coefficient "Action Unit 6" which has been called "cheek raiser". Using the earlier version of Kinect sensor only 6 action units could be detected, whereas 17 action units (AUs) can be tracked with the new Kinect v2 and the high definition face tracking API. Enumeration of action units tracked with Kinect v1 and Kinect v2 sensors is shown in Fig. 1.

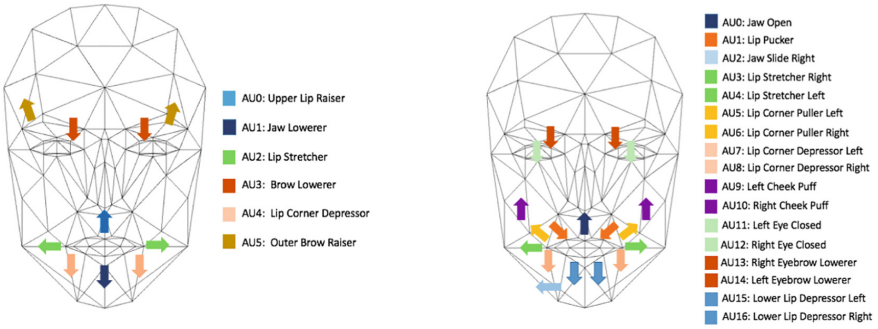


Fig. 1. Visualization of action units (AUs) extracted from Kinect v1 and v2, respectively. Color-coded labels of AUs indicate the position of that specific muscle. The arrow signs are used to illustrate the muscle movement. (Color figure online)

Over the past few decades facial and emotion recognition have attracted a great deal of research interest. Existing research efforts can be classified as image-based, video-based, and 3D surface-based methods [3]. While 3D facial recognition techniques were investigated in the literature to some extent, most of the studies have focused on two-dimensional (2D) algorithms which are computationally expensive [4–8].

Mao et al. [9] proposed a real-time EFRE method, in which, six action units (AUs) and 45 feature point positions (FPP) are used as features. Using these two features respectively, the classification of emotions has been done using support vector machine (SVM) classifiers and the recognition results of 30 consecutive frames are fused by the fusion algorithm based on improved emotional profiles (IEPs). In 2013, Youssef et al. [10] constructed a dataset containing 3D data for 14 different persons performing the 6 basic facial expressions. SVM and k-NN are used to classify emotions. They have achieved 38.8% (SVM) and 34.0% (k-NN) accuracy for individuals who did not participate in training of the classifiers, whereas observed 78.6% (SVM) and 81.8% (k-NN) accuracy levels considering individuals who did participate in training. Zhang et al. [11] collected the 3D facial points recorded by Kinect from 52 subjects (34 female and 18 male). The best accuracy reached was 80% for three emotions in only female data

with decision tree classification. Lastly, in 2017, Tarnowski et al. [12] performed emotion classification using 6 action units tracked by Kinect v1 sensor as features to neural network. They have constructed the dataset with six men performing 7 emotional states, total of 256 facial expressions.

In this paper, we propose a facial emotion recognition approach based on several action units (AUs) tracked by a Kinect v2 sensor to recognize six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise) and neutral. Classification is performed through artificial neural networks (ANNs) where inputs of the network are features obtained from a 3D depth camera. Microsoft Kinect for Windows sensor v2 and high definition face tracking SDK have been used for tracking the facial action units (AUs) [13]. Unlike the existing method [12] where 6 features were employed for classification, we use 17 action units (AUs) derived from FACS system as features. As it can be seen from Fig. 1, these newly added 11 action units are mainly located in the lower face and thus provide a better representation for emotions that involve movement of muscles around mouth area. It should be noted that the new set of features plays a distinctive role in classifying complex human emotions. Several experiments are conducted on a homemade dataset which consists of 3 male and 3 female subjects. Classification results show the potential of our proposed method.

This paper is organized as follows: Sect. 2 describes our method in detail. Classification results are presented and discussed in Sect. 3. The paper is finalized with a conclusion in Sect. 4.

2 Method

2.1 Dataset and Experimental Setup

Six volunteers (3 women and 3 men) took part in data generation process. The participants were seated at a distance of two meters away from the Kinect sensor. Two sessions were held in which subjects mimicked all seven emotional states: neutral, joy, surprise, anger, sadness, fear and disgust. Each participant took 10 second breaks between emotional states and performed each emotion for a minute. Sample images from our dataset is shown in Fig. 2. For each session, 5 peak frames have been chosen per participant per emotion. As a result, 70 frames (5 peak frames $\times$ 2 sessions $\times$ 7 emotions) were collected for each subject. Overall dataset consisted of 420 (70 frames $\times$ 6 participants) facial expressions.

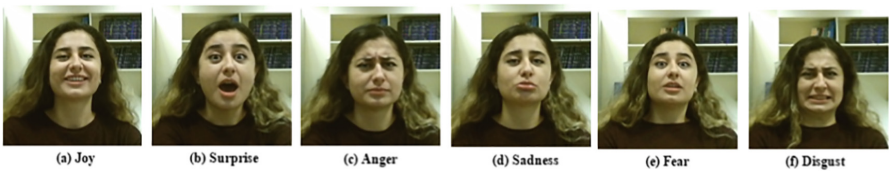


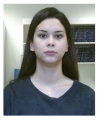
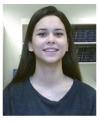
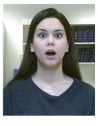
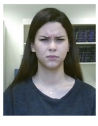
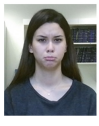
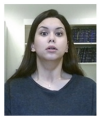
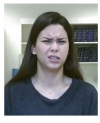
Fig. 2. Example of facial expressions of one subject from dataset.

The high definition face tracking API developed by Microsoft, was employed to extract seventeen action units (AUs). 14 out of 17 AUs are expressed as a numeric weight varying between 0 and 1 whereas the remaining 3, Jaw Slide Right, Right Eyebrow Lowerer, and Left Eyebrow Lowerer, vary between −1 and +1, representing the displacement from a neutral AU expression. The action unit features coming from each frame can be written in the vector form:

$$a = (AU_0, AU_1, AU_2 ... AU_{16}) \tag{1}$$

Then, these action units were used as features in the classification process. Table 1 illustrates the exemplary values of action units for different emotional states (neutral, joy, surprise, anger, sadness, fear and disgust) of one participant.

Table 1. Numerical values of action units corresponding to various facial expressions.

	neutral	joy	surprise	anger	sadness	fear	disgust
							
AU0	0.045	0.160	0.337	0.088	0.091	0.229	0.197
AU1	0.231	0.000	0.347	0.216	0.392	0.246	0.289
AU2	0.057	-0.019	0.018	-0.055	0.001	-0.067	-0.037
AU3	0.133	0.204	0.010	0.101	0.052	0.015	0.022
AU4	0.115	0.161	0.032	0.101	0.151	0.188	0.030
AU5	0.000	0.820	0.023	0.020	0.010	0.024	0.084
AU6	0.015	0.903	0.023	0.025	0.014	0.002	0.013
AU7	0.040	0.056	0.174	0.118	0.421	0.241	0.116
AU8	0.044	0.080	0.134	0.036	0.437	0.277	0.193
AU9	0.034	0.022	0.019	0.026	0.035	0.020	0.022
AU10	0.031	0.010	0.016	0.018	0.025	0.021	0.024
AU11	0.201	0.179	0.047	0.132	0.055	0.083	0.304
AU12	0.216	0.127	0.033	0.044	0.046	0.034	0.366
AU13	0.340	-0.049	-0.246	0.156	-0.118	-0.201	0.117
AU14	0.371	0.063	-0.221	0.376	-0.025	-0.085	0.258
AU15	0.000	0.550	0.036	0.045	0.017	0.001	0.137
AU16	0.000	0.482	0.026	0.044	0.018	0.005	0.114

2.2 Classification

Having extracted the action units from Kinect sensor, we provided this features to our classifier as input. The flowchart of our proposed method can be seen in Fig. 3.

We have used a neural network classifier with one hidden layer to classify emotional states. 10 neurons and sigmoid action function have been used in hidden layer. Input layer consisted of seventeen action units (AUs). The output was one of the seven emotional states. We trained the neural network using scaled conjugate gradient backpropagation algorithm. The structure of the neural network can be seen in Fig. 4.

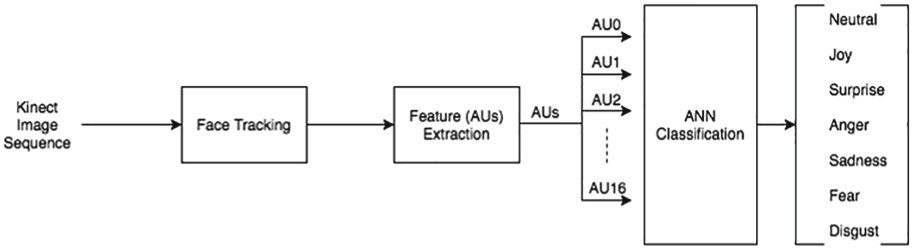


Fig. 3. Schematic representation of the methodology.

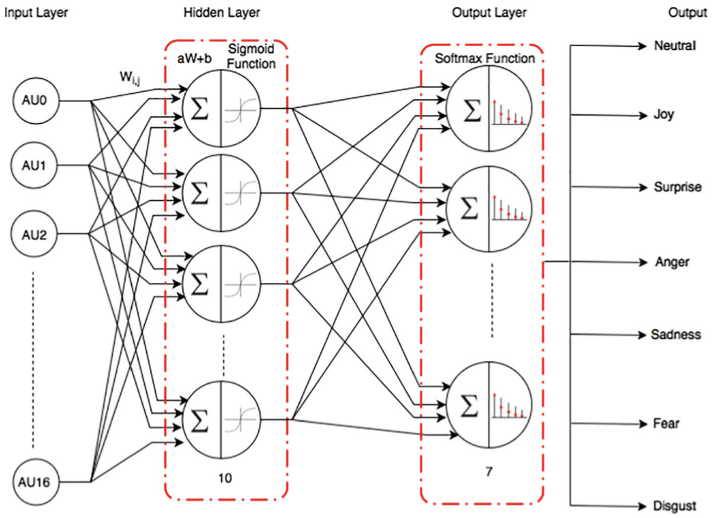


Fig. 4. The neural network classifier.

3 Experimental Results

We first tested our method for subject dependent case in which all the data were randomly divided into training, testing and the validation parts. Training part consisted of 70% of our overall data (294 samples) whereas testing and validation parts were 15% each (63 samples). The validation data is used to measure model generalization, and to terminate training when generalization stops improving. Testing samples which have no effect on training is used to measure network performance. In subject dependent case, we have tested our approach for the same volunteers took part in our data generation process. The network was trained 50 times using scaled conjugate gradient backpropagation. The average classifier accuracy was 95.8% for testing set and 96.2% for validation set (Table 2). AQ2

Afterwards, we evaluated our approach through training our network with samples collected from 5 volunteers and testing it with 6th volunteer who was

Table 2. Classification performances.

Training data	Validation data	Test data	Test accuracy (%)
70% of dataset	15% of dataset	15% of dataset	95.80

Table 3. Classification performances.

Test subject	Classification accuracy (%)
1	74.3
2	57.1
3	80.6
4	52.8
5	66.2
6	71.2
Average	67.03

Fig. 5. The neural network classifier.

not part of the training data. This experiment is repeated for each subject. The network's average performance was 67.03% for classifying images which were not part of its training dataset (Table 3).

It has been known that skin color, facial hair and gender play a significant role in the quality of emotion recognition system. To examine the effect of gender in our classification approach, we performed an additional test. We divided our dataset into half and used the samples collected from men as training data. We then tested the network with remaining samples coming from our three female volunteers. For this gender-based test, we obtained 56% classification accuracy. Samples from both training and test set are shown in the Fig. 5, along with the classification accuracy.

Output Class \ Target Class	neutral	joy	surprise	anger	sadness	fear	disgust
neutral	10	0	0	0	0	0	1
joy	0	8	0	0	0	0	0
surprise	0	0	13	0	0	0	0
anger	0	0	0	7	0	0	0
sadness	0	0	0	0	8	0	0
fear	0	0	0	0	0	8	1
disgust	0	0	0	0	0	0	7

Fig. 6. Test confusion matrix.

Confusion matrices are shown in Fig. 6 to determine easiest and most difficult emotions in term of classification. Distribution of 63 test samples into the classes neutral, joy, surprise, anger, sadness, fear and disgust can be seen from Fig. 6. According to the confusion matrix, disgust, neutral and fear are most likely to be misclassified. Lighting of the environment, head orientation of subject, distance from the sensor, and distinctive characteristics of human expressions are the main challenges for which Kinect could identify the feature points.

4 Conclusion

In this paper, we have presented our facial emotion recognition approach to classify seven different emotional states. We have created our own dataset of 420 samples collected from six volunteers with different gender and ethnicity. Each sample is labeled as neutral, joy, sadness, anger, surprise, fear or disgust. We have achieved average accuracy of 95.8% for subject dependent case and 67.03% accuracy for images of a different subject which did not take part in training. Moreover, we have examined the impact of gender by training the network with samples collected from male volunteers and testing it with female subjects. In this case, we have obtained 56% classification accuracy which is significantly lower compared to previous cases. In the future work, to eliminate the effect of gender, we plan to work on a fusion algorithm based on action units (AUs) and feature point positions (FPPs) and use principal component analysis (PCA) for feature selection process. Furthermore, we will expand our dataset by adding people with different ages.

References

1. Mehrabian, A.: Communication without words. In: Mortensen, C.D. (ed.) Communication Theory, pp. 193–200. Transaction Publishers, New Brunswick (2008)
2. Ekman, P., Friesen, W.: Facial Action Coding System. Consulting Psychologists Press, Stanford University, Palo Alto (1977)
3. Wang, P., Barrett, F., Martin, E., Milonova, M., Gur, R.E., Gur, R.C., Kohler, C., Verma, R.: Automated video-based facial expression analysis of neuropsychiatric disorders. J. Neurosci. Methods **168**(1), 224–238 (2008)
4. Tong, Y., Chen, R., Cheng, Y.: Facial expression recognition algorithm using LGC based on horizontal and diagonal prior principle. Optik-Int. J. Light Electron. Optics **125**, 4186–4189 (2014)
5. Jabid, T., Kabir, M.H., Chae, O.: Facial expression recognition using local directional pattern (IDP). In: 17th IEEE International Conference on Image Processing (ICIP). IEEE (2010)
6. Guo, Y., Tian, Y., Gao, X., Zhang, X.: Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method. In: International Joint Conference on Neural Networks (IJCNN). IEEE (2014)
7. Gizatdinova, Y., Surakka, V., Zhao, G., Makinen, E., Raisamo, R.: Facial expression classification based on local spatiotemporal edge and texture descriptors. In: Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research. ACM (2010)
8. Kabir, M.H., Jabid, T., Chae, O.: Local directional pattern variance (IDPV): a robust feature descriptor for facial expression recognition. Int. Arab. J. Inf. Technol. (2012) AQ3
9. Mao, Q., Pan, X., Zhan, Y., Shen, X.: Using Kinect for real-time emotion recognition via facial expressions. Front. Inf. Technol. Electron. Eng. **16**(4), 272–282 (2015)
10. Youssef, A.E., Aly, S.F., Ibrahim, A.S., Abbott, A.L.: Auto-optimized multimodal expression recognition framework using 3D kinect data for ASD therapeutic aid. Int. J. Model. Optim. (2013)
11. Zhang, Z., Cui, L., Liu, X., Zhu, T.: Emotion detection using Kinect 3D facial points. In: International Conference on Web Intelligence (2016)
12. Tarnowski, P., Kolodziej, M., Majkowkski, A., Rak, R.: Emotion recognition using facial expressions. In: International Conference on Computer Science, ICCS (2017)
13. Microsoft SDK for Face Tracking. http://msdn.microsoft.com/enus/library/jj130970.aspx

Author Queries

Chapter 24

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	
AQ2	Please check and confirm if the inserted citation of Tables 2 and 3 are correct. If not, please suggest an alternate citation.	
AQ3	Please provide volume number and page range for Refs. [8, 10].	