Chapter 7
Facial Emotion Recognition Based on Cascade of Neural Networks

Elżbieta Kukla and Paweł Nowak

Abstract. The chapter presents a method that uses the cascade of neural networks for facial expression recognition. As an input the algorithm receives a normalized image of a face and returns the emotion that the face expresses. To determine the best classifiers for recognizing particular emotions one- and multilayered networks were tested. Experiments covered different resolutions of the images presenting faces as well as the images including regions of mouths and eyes. On the basis of the tests results a cascade of the neural networks was proposed. The cascade recognizes six basic emotions and neutral expression.

7.1 Introduction

Human emotions expressed by mimics play fundamental role in everyday communication. Nonverbal information conveyed during conversation permits akes it possible to properly interpret and understand meaning of an utterance as well as the intentions of an interlocutor. Human brain recognizes mimics in a split second. For this reason emotion recognition became an important element of “natural” dialog between user and computer system.

Psychologists separated six basic emotions that are universal and occur in every culture. These are: happiness, sadness, fear, anger, disgust and surprise [2]. Although, the task of facial emotion recognition seems to be simple and intuitive for most of the people it is not easy for computer systems. One of the reasons is that every emotion can be expressed in many different ways, e.g. for fear there are about 60 various face expressions that have some common features [1]. For remaining emotions this number is similar. So, it is necessary to distinguish them somehow even if the differences are sometimes really subtle.

Elżbieta Kukla · Paweł Nowak
Institute of Informatics, Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: {elzbieta.kukla,pawel.nowak}@pwr.edu.pl

A. Zgrzywa et al. (eds.), *New Research in Multimedia and Internet Systems*,
Advances in Intelligent Systems and Computing 314, DOI: 10.1007/978-3-319-10383-9_7

The chapter presents an approach to solve the problem of facial emotion recognition using a cascade of the neural networks trained to recognize individual emotions. The solution was tested on three different sets of photographs. Results obtained confirm intuitions and earlier observations made by other authors.

The chapter is organized as follows. Second part presents facial emotion recognition problem and some of the solutions reported in the literature. Part three describes in detail the way a cascade of neural networks is constructed as well as the tests that were carried out to verify its performance. Last section of this part reports and discusses results achieved for different sets of photos. Fourth part contains general conclusions and presents plans for future works.

7.2 Facial Emotion Recognition

Modern facial emotion recognition systems receive as input both pictures and videos. They all have a similar structure (Fig. 7.1) and consist of three main modules [11]: face capturing, extraction of the face features and emotion recognition.

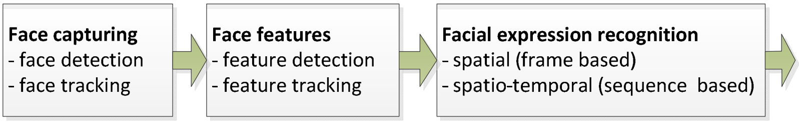

Fig. 7.1 General structure of the systems analyzing face expression [11]

Face capturing includes face detection and tracking that is connected with estimation of the face position. In the space of years many methods and algorithms concerning face detection have been developed. Their exhaustive review is presented among others by Yang, Kriegman and Ahuja in [13]. Separate category of problems is face tracking. In this field position [10] and especially chapter “Image and Video Processing Tools for HCI” gives comprehensive literature review.

Extraction of face features comprises feature detection and tracking. Generally, the approaches used to solve this problem can be categorized into local feature detection, global model detection and hybrid systems [10].

Facial emotion recognition systems are based on two approaches. The first of them classifies facial expression to one of the categories representing emotions. The second approach relies on identifying and measuring facial muscles motions (FACS) that delivers more detailed information about facial expressions.

Classification approach uses vectors of features that are characteristic for face appearance as well as for their movement. Features are often gathered during the process of their detecting and tracing. Depending on the solution categorization can be based on spatial- (referring to a single frame) or spatio-temporal (referring to a sequence of video frames) classifiers [4]. While spatial classifiers commonly use neural networks, spatio-temporal ones utilize Hidden Markov Models [9], [7].

The facial Action Coding System (FACS) [2] is based on human observations. It detects changes in particular parts of face [11]. FACS contains all the Action Units (AUs) of a face that cause facial movement. Part of them is closely connected with specific face muscle. The majority of Action Units permit both symmetric and asymmetric coding of facial actions, e.g. closing one eye. For AU that can be more or less intensive it is possible to use three- or five-level scales. Individual AU may form more complex actions that can be found in real situations. FACS system alone is designed for the actions descriptions and does not offer the possibilities of an inference about the emotions expressed. For this purpose distinct systems were developed. One of the examples is EMFACS (Emotional Facial Action System) that uses combinations of AC from FACS to determine the emotions expressed [Friesen 1983]. The other instance is FACSAID (Facial Action Coding System Affect Interpretation Dictionary) [3] that describes the meanings of particular behaviors represented by the combinations of AU and in this way gives the systems of facial expression recognition an ability of an interpretation of their results. The experts (psychologists) have assigned an emotion to every combination of AU.

7.3 Cascade of Neural Networks Applied to Facial Expressions Recognition

The experiment, presented in the subsequent part of this section, investigates in what way a cascade of neural networks will recognize six basic facial emotion expressions depending on image size, color, lighting and focus. This kind of recognition has been proposed by Golomb [5] and used for the first time in a system for female/male sex recognition.

At the first stage, various (one- and multilayer) neural networks were considered and compared to find out the six the most suitable networks for facial emotion recognition, one network for a particular emotion. Basing on this research, at the second stage, a cascade of neural networks has been proposed and tested to determine the influence of image representation on classification results.

All the experiments used three sets of photographs presenting six basic emotions: fear, anger, disgust, happiness, sadness, surprise and additionally neutral face expression. The photographs originated from The Karolinska Directed Emotional Faces (KDEF) data base [8], John Kanade [6] data base and a set of photographs of the students gathered by the authors.

The images chosen from three databases mentioned above were preprocessed to obtain the pictures that composed an input to cascade of classifiers.

7.3.1 Sets of Photographs

The Karolinska Directed Emotional Faces (KDEF) data base is one of the biggest set of photographs available in Internet. KDEF contains images of 70 faces: 35 men and 35 women that are 20–30 years old and none of them wears bread or glasses. The photographs present full-faces that express six basic emotions: fear, anger, disgust, happiness, sadness and surprise. All the pictures are colorful and have the same dimensions 562x762 pixels. The data base comprises two series of pictures that present actors expressing the same emotion twice (Fig. 7.2). One of the series was used for training classifier and the other – for testing it.

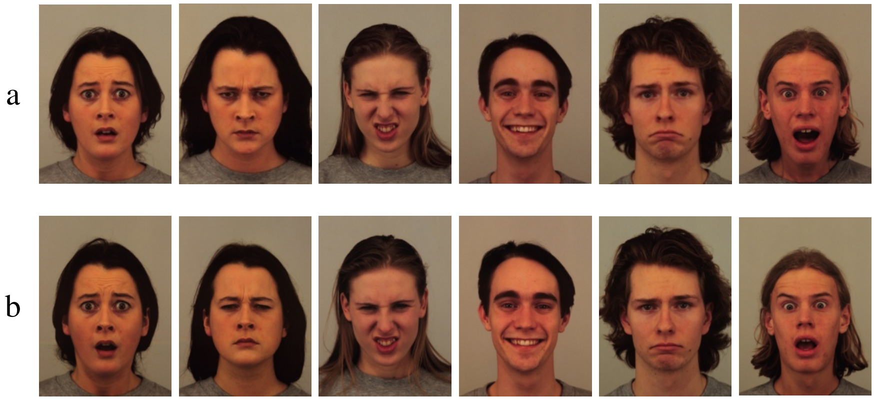

Fig. 7.2 An example of KDEF subsets used for (a) training and (b) testing classifier [8]

The John Kanade data base of photographs is often used in scientific research related to facial emotion recognition. It contains grey scale, 640x490 pixels images. For the purpose of this work 16 photographs of different persons were chosen for every emotion to be recognized. Fig. 7.3 presents exemplary pictures from Kanade's data base.

Third set of photographs consists of the color pictures in size of 800x600 pixels. The photo present six basic emotions expressed by students, from 14 to 20 snaps for one emotion. The set was collected by one of the authors.

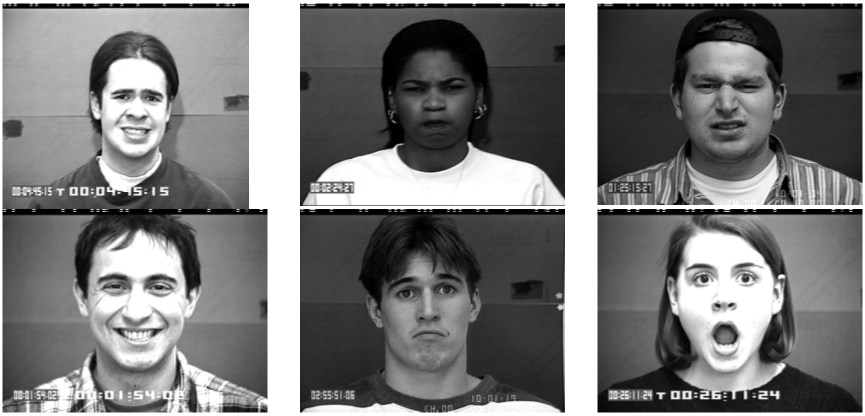

Fig. 7.3 Exemplary test photographs from John Kanade database [6]

7.3.2 Selection of Neural Networks for Emotion Recognition

Main idea of the research reported in this chapter was the application of a cascade of the classifiers to facial emotion recognition. Each of the classifiers was a single neural network or a configuration of two neural networks that was able to recognize the most fitting, single emotion. Configuration of two neural networks was used when an emotion was expressed in two different ways by different “actors”. Then, the final result of the recognition was maximum of the results given by both of the networks. Such a situation occurred in case of anger and surprise that were expressed with closed or opened mouth, disgust that was expressed with closed mouth or opened mouth and visible teeth and sadness that was expressed by lowered mouth corners or in any other way. For fear and happiness recognition the classifier consisted of a single neural network.

The elements of the cascade were determined experimentally. Two kinds of neural networks were considered: a network consisted of one perceptron (PERC) and a three-layered network (MLP) with one exit neuron and five hidden neurons in the intermediate layer. The numbers of the networks entries depended on the size of the images that were recognized. For the purpose of the experiment, the authors took under consideration three kinds of pictures. Two of them consisted of 38x38 pixels (1444 entries) and 50x50 pixels (2500 entries). Third input image composed of two regions including mouth (30x30 pixels) and eyes (15x16 pixels), what gives total 1700 entries.

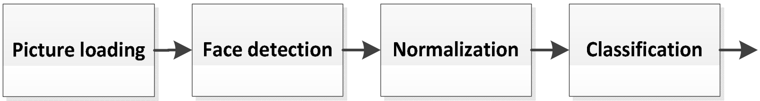

Fig. 7.4 Schema of the experiment

The six networks: 1444PERC, 1444MLP, 2500PERC, 2500MLP, 1700PRC and 1700MLP were examined according to the schema presented on Fig. 7.4. First, the images were loaded to test system. Then, Viola and Jones [12] algorithm detected faces on pictures. On every image a region containing face was reduced to the size of 300x300 pixels. If original region was smaller than 300x300 pixels it was enlarged to this dimension. Normalization that was accomplished at next step consisted in pruning peripheries, i.e. cutting background that occupies about 1/8 of image, and reducing its size to the three dimensions mentioned above. During this stage the colorful pictures were converted into the grey scale images and their histograms were smoothed. The normalized pictures were converted into the vectors that composed entrances for the networks bequeathed. The pixels of the images were the coefficients of the equivalent vectors. The last step was classification that aimed at appointing the best neural network for recognizing given emotion. All the networks were first trained. For training multilayered networks backward propagation algorithm was used with learning coefficient 0.01 and momentum 0.7

until max error level 0.00005 or number of iteration equal to 10000. Perceptron networks were trained with learning coefficient 0.2 and max error level 0.00005. Every network was tested then three times using KDEF database, Kanade database and the third set of photographs made by the author. Results obtained in particular tests are presented in Tables 7.1, 7.2 and 7.3. Symbols TP and FP means true positive and false positive recognition respectively. The best results of recognition for each emotion are bolded.

Table 7.1 Results of the tests for KDEF database

		Fear	Anger	Disgust	Happiness	Sadness	Surprise
1444PERC	TP	0.57	0.53	**0.84**	0.94	0.73	0.81
	FP	0.03	0.02	0.06	0.01	0.10	0.02
1444MLP	TP	0.46	0.57	0.77	0.96	0.76	0.86
	FP	0.01	0.01	0.03	0.02	0.06	0.01
2500PERC	TP	0.54	**0.61**	0.76	**0.97**	0.70	0.76
	FP	0.04	0.02	0.05	0.02	0.08	0.01
2500MLP	TP	0.49	0.56	0.76	0.96	**0.83**	**0.94**
	FP	0.01	0.01	0.04	0.02	0.09	0.02
1700PERC	TP	**0.63**	0.59	0.79	**0.97**	0.60	0.81
	FP	0.03	0.02	0.04	0.02	0.05	0.02
1700MLP	TP	0.49	0.60	0.77	0.94	0.67	0.89
	FP	0.01	0.02	0.04	0.02	0.04	0.02

The KDEF database contains two subsets of the pictures that present actors expressing the same emotion twice and therefore one of them was used for networks training, the other – for testing them. In this case (Table 7.1), happiness expressed by a smile achieved the best results of recognition 0.97 for 2500PERC and 1700PERC networks. Next was surprise with recognition 0,94 and 2500MLP network. Disgust gained recognition 0,84 for 1444PERC network. Only a little bit worse results 0,83 were obtained for sadness and 2500MLP network. Fear was recognized in 63% by 1700PERC network. The worst effects of recognition 0.61 were achieved for anger and 2500PERC network. False positive results in all the tests are not greater than 0.1. This value was obtained for sadness and disgust. Generally, better results were gained for perceptron networks. Multilayered network was better only in the case of surprise recognition.

The tests based on Kanade database and set of author's photographs were constructed somewhat differently. In these two cases the sets contained one unique photo per actor and per emotion. So, for every emotion two subsets were isolated basing on the Kanade database and two subsets from author's photo set respectively. Every subset contained photos of different persons expressing the same emotion.

Table 7.2 Results of the tests for Kanade database

		Fear	Anger	Disgust	Happiness	Sadness	Surprise
1444PERC	TP	0.00	0.19	**0.38**	0.75	**0.25**	**0.81**
	FP	0.02	0.06	0.04	0.03	0.23	0.09
1444MLP	TP	0.00	0.19	0.13	**0.81**	0.06	0.75
	FP	0.00	0.03	0.00	0.07	0.06	0.05
2500PERC	TP	0.00	**0.50**	0.25	**0.81**	0.31	0.75
	FP	0.02	0.11	0.03	0.07	0.24	0.03
2500MLP	TP	0.00	0.19	0.19	**0.81**	0.06	0.75
	FP	0.00	0.05	0.01	0.07	0.06	0.10
1700PERC	TP	**0.06**	0.00	0.19	0.75	0.00	0.69
	FP	0.03	0.01	0.02	0.08	0.06	0.06
1700MLP	TP	**0.06**	0.00	0.19	**0.81**	0.00	0.69
	FP	0.00	0.01	0.01	0.08	0.05	0.05

For Kanade database (Table 7.2) the best results of recognition 0.81 were achieved for happiness and surprise. The best recognitions of happiness was achieved by 1444MLP, 2500PERC and 2500MLP networks. For surprise the best identification effects were obtained by 1444PERC network. Remaining recognition results presented as follows: for anger 0.50 and 2500PERC network, disgust 0.38 and sadness 0.25 by 1444PERC network, and at last fear 0.06 by 1700PERC and 1700MPL networks. It is easy to notice that the results achieved in this test were inferior than in the case of photos from KDEF database. Also false positive results were inferior and reached even up to 0.24.

Table 7.3 Results of the tests for author's set of photo

		Fear	Anger	Disgust	Happiness	Sadness	Surprise
1444PERC	TP	0.00	0.06	**0.15**	0.48	**0.29**	0.50
	FP	0.01	0.19	0.12	0.02	0.28	01.
1444MLP	TP	0.00	0.19	0.00	0.52	**0.29**	0.56
	FP	0.00	0.13	0.01	0.04	0.11	0.03
2500PERC	TP	0.00	**0.44**	0.08	**0.57**	**0.29**	0.28
	FP	0.00	0.19	0.11	0.05	0.24	0.04
2500MLP	TP	0.00	0.06	0.00	**0.57**	**0.29**	0.56
	FP	0.00	0.14	0.05	0.05	0.12	0.14
1700PERC	TP	0.00	0.00	0.08	0.52	0.21	**0.67**
	FP	0.01	0.11	0.09	0.06	0.10	0.06
1700MLP	TP	0.00	0.06	0.00	0.48	0.21	0.67
	FP	0.00	0.11	0,02	0.03	0.04	0.04

The tests performed on author's photos set (Table 7.3) returned the best results for surprise (0.67) and 1700PERC network. Next was happiness with recognition ratio 0.57 gained by 2500MLP and 2500 network. Anger was recognized the best (0.44) by 2500PERC network. Recognition of sadness reached value 0.29 by 1444PERC, 1444MLP, 2500PERC and 2500MLP networks. Disgust was recognized by 1444PERC network at level of 0.15. The worst result of recognition (0.00) was achieved for fear. False positive results reached maximum value 0.28 for sadness.

The tests revealed that the best recognition results were obtained for the photos from the KDEF database. Therefore, they served to determine the elements of a cascade of neural networks. Figure 7.5 presents the cascade where every emotion is recognized by a single classifier (the best for this emotion in KDEF test). The neural networks in the cascade are arranged in ascending order of false positive results.

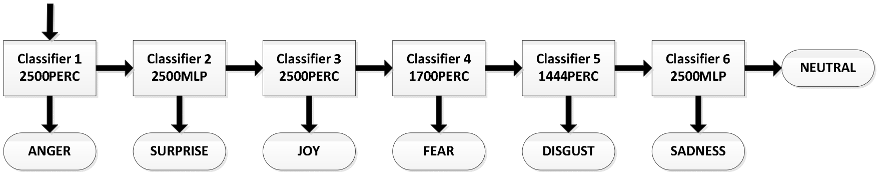

Fig. 7.5 The cascade of classifiers for facial emotion recognition

Every succeeding classifier assigns an input picture to one of the emotion class, for example the first classifier recognizes anger and if it does not identify this emotion the picture passes to the next classifier. The procedure concludes whenever the picture is classified to one of the classes representing six basic emotions. If none of the classifiers recognizes the picture then it is supposed that face at this picture presents neutral expression.

Additionally, particular networks composing the cascade were visualized. The images arisen as the visualization effects permitted to discover the features of the faces that decided about the results of the classifications.

Neural networks containing one perceptron and recognizing anger were trained to identify two mimic expressions. For anger expressed by closed mouth the pixels from the regions of eyebrows (their middle part), lips and chin had the greatest influence on the recognition results. For picture where anger is expressed by opened mouth the most important appeared the pixels that represented eyebrows, open mouth with visible teeth as well as the lines running from nose to mouth corners.

In the multilayered network that recognized surprise at the pictures with widely opened mouth it was possible to distinguish several groups of "deciding" neurons. One of them are concentrated on the region of mouth, the others – on the region of eyebrows, nose and middle part of brow. Surprise expressed by faces with closed mouth was recognized the best by multilayered neural network with the most important neurons focused around eyebrows and nose.

The perceptron network for happiness recognition took into account mainly the region of mouth with visible corners and wrinkles accompanying smile. In the case of fear recognition the most important are terrified eyes and opened mouth with and opened mouth with lines running from nose to mouth corners. The visualization of neutron that recognized disgust expressed by closed mouth revealed the most important regions around eyebrows and beneath nose. Disgust expressed by open mouth was recognized the best basing on wrinkles around nose, teeth and wrinkled eyebrows. In sadness recognition the most important neutrons from multilayered network are focused around lowered mouth corners and region of eyebrows.

7.3.3 Studies of the Facial Emotion Recognition by the Cascade of Classifiers

Facial emotion recognition based on the cascade of classifiers was tested using three sets of pictures described in section 7.3.2. All the experiments were carried out according to the schema presented at Fig. 7.4.

Results obtained for KDEF database are presented in Table 7.4. The best effect of recognition was achieved for happiness (0.97) and the worst for anger (0.51). Fear was often classified as surprise (0.11) what might be caused by widely opened eyes and half-opened mouth that were characteristic both for fear and surprise. Similarly, anger was classified as disgust (0.09) with respect to wrinkled eyebrows and as sadness (0.1) when lowered mouth corners were taken into account. Sadness in turn was confused with disgust (0.1) what might be caused by the fact that some actors wrinkled their noses and eyebrows.

Table 7.4 Emotion recognition for KDEF database

In\Out	Fear	Anger	Disgust	Happiness	Sadness	Surprise	Neutral
Fear	**0.54**	0.00	0.00	0.03	0.06	0.11	0.26
Anger	0.00	**0.51**	0.09	0.00	0.10	0.03	0.26
Disgust	0.00	0.00	**0.79**	0.06	0.06	0.00	0.10
Happiness	0.00	0.00	0.00	**0.97**	0.01	0.00	0.01
Sadness	0.01	0.00	0.10	0.00	**0.76**	0.00	0.13
Surprise	0.03	0.00	0.00	0.00	0.00	**0.94**	0.03
Neutral	0.01	0.07	0.03	0.00	0.00	0.04	**0.84**

The next test referred to identification of facial expressions presented at the photos from Kanade database. Recognition results achieved by the cascade in this case are presented in Table 7.5. Like in the previous experiment the best recognized emotion was happiness (0.81). The worst results were obtained for sadness (0.00) – none of pictures presenting sad face was identified. Sadness was relatively often (0.19) confused with anger and surprise what could be caused by significant differences between the photos used in training phase and the photos from Kanade's database.

Table 7.5 Emotion recognition for Kanade database

In\Out	Fear	Anger	Disgust	Happiness	Sadness	Surprise	Neutral
Fear	**0.06**	0.00	0.06	0.44	0.00	0.06	0.38
Anger	0.00	**0.50**	0.00	0.00	0.06	0.06	0.38
Disgust	0.00	0.38	**0.25**	0.00	0.00	0.06	0.31
Happiness	0.06	0.06	0.00	**0.81**	0.06	0.00	0.00
Sadness	0.00	0.19	0.00	0.00	**0.00**	0.19	0.63
Surprise	0.00	0.00	0.00	0.00	0.00	**0.75**	0.25
Neutral	0.06	0.06	0.00	0.00	0.06	0.06	**0.75**

The last experiment dealt with recognition of the emotions expressed by the faces from the author's collection of the photos. Table 7.6 presents the results achieved in this case. Fear was not recognized at all (0.00). Identification of the remaining emotions varied from 0.14 for sadness up to 0.57 for happiness.

Generally the results obtained in this test were worse than in previous tests. This can result from the fact that people presented at the pictures were not professional actors and the emotions presented by them were expressed intuitively. Additionally, photos from the third set were taken in unprofessional manner with respect to the equipment, conditions, lighting etc.

Figure 7.6 presents positive results of facial emotion recognition based on the cascade of neural networks for three sets of photographs presented in details and discussed in previous paragraphs.

Table 7.6 Emotion recognition for the author's set of the photographs

In\Out	Fear	Anger	Disgust	Happiness	Sadness	Surprise	Neutral
Fear	**0.00**	0.07	0.13	0.33	0.00	0.00	0.47
Anger	0.00	**0.44**	0.00	0.00	0.06	0.06	0.44
Disgust	0.00	0.23	**0.08**	0.00	0.15	0.00	0.54
Happiness	0.05	0.00	0.05	**0.57**	0.05	0.00	0.29
Sadness	0.00	0.07	0.00	0.00	**0.14**	0.14	0.64
Surprise	0.00	0.00	0.00	0.00	0.06	**0.56**	0.39
Neutral	0.00	0.06	0.06	0.00	0.12	0.35	**0.41**

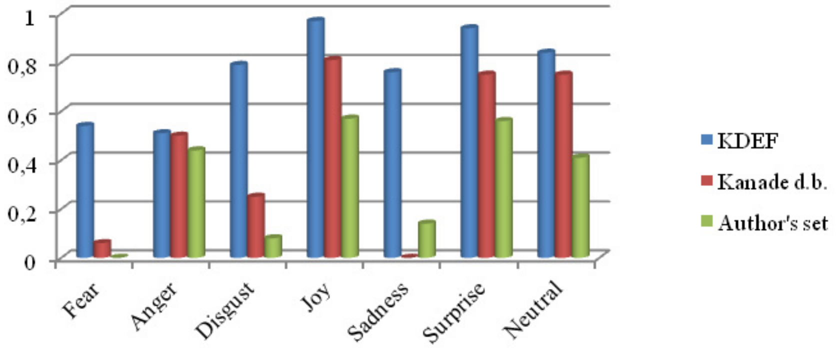

Fig. 7.6 Positive results of six basic recognition for three sets of photos

7.4 Conclusions and Future Works

The cascade of the neural networks developed in these studies returned satisfactory results (on average 76% for KDEF database) but it did not achieved the efficiency close to that of recent facial emotion recognition systems (>95%). The results obtained are different for particular emotions. The worst recognized emotion is fear, the best is happiness and surprise. It is worth to notice that exact comparison of the cascade with the other systems described in the literature is difficult because the authors could not have an access to the photos used in testing them.

Future investigations will concentrate on separate recognition of the two parts of faces that are the most important in emotion identification, i.e. regions of eyes, eyebrows and forehead (upper parts) and regions of mouth, nose and chin (lower parts). Particular networks could be trained and used to recognize separately particular action units from FACS coding system. Final recognition results is this case could be a fusion of the results obtained by individual classifiers.

References

1. Ekman, P.: Facial expression and emotion. American Psychologist 48(4), 384, 384–392 (1993)
2. Ekman, P., Friesen, W., Hager, J.: Facial action coding system. Consulting Psychologists Press, Palo Alto (1978)
3. Ekman, P., Hager, J., Rosenberg, E.: FACSAID: A computer database for predicting affective phenomena from facial movement (2003), `http://face-and-emotion.com/dataface/facsaid/description.jsp`, `http://face-and-emotion.com/dataface/nsfrept/psychology.html` (visited April 4, 2014)
4. Fasel, B., Luettin, J.: Automatic facial expression analysis: A survey. Pattern Recognition Society 36(1), 259–275 (2003)
5. Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: Sexnet: A neural net identifies sex from human faces. In: Lippman, R.P., Moody, J., Touretzky, D.S. (eds.) NIPS, vol. 3, pp. 572–577. Morgan Kaufmann, San Francisco (1991)
6. Kanade, T., Cohn, J., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the Fourth IEEE International Conference on Automatic Face Gesture Recognition (FG 2000), Grenoble, France, pp. 46–53 (2000)
7. Lien, J., Kanade, T., Cohn, J., Li, C.: Detection, tracking and classification of action units in facial expressions. Robotics and Autonomous Systems 31(3), 131–146 (2000)
8. Lundqvist, D., Flykt, A., Öhman, A.: The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, pp. 91–630 (1998)
9. Pardàs, M., Bonafonte, A.: Facial animation parameters extraction and expression recognition using Hidden Markov Models. Signal Processing: Image Communication 17(9), 675–688 (2002)
10. Thiran, J.P., Marques, F., Bourlard, H. (eds.): Multimodal Signal Processing: Theory and Applications for Human-Computer Interaction. Elsevier, San Diego (2010)

11. Tian, Y., Kanade, T., Cohn, J.: Facial expression analysis. In: Handbook of Face Recognition, pp. 247–276 (2005)
12. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, pp. I-511–I-518 (2001)
13. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1), 34–58 (2002)