Recognizing Emotions from Facial Expressions Using Neural Network

Isidoros Perikos, Epaminondas Ziakopoulos, and Ioannis Hatzilygeroudis

School of Engineering
Department of Computer Engineering & Informatics
University of Patras
26500 Patras, Hellas, Greece
{perikos,ziakopoul,ihatz}@ceid.upatras.gr

Abstract. Recognizing the emotional state of a human from his/her facial gestures is a very challenging task with wide ranging applications in everyday life. In this paper, we present an emotion detection system developed to automatically recognize basic emotional states from human facial expressions. The system initially analyzes the facial image, locates and measures distinctive human facial deformations such as eyes, eyebrows and mouth and extracts the proper features. Then, a multilayer neural network is used for the classification of the facial expression to the proper emotional states. The system was evaluated on images of human faces from the JAFFE database and the results gathered indicate quite satisfactory performance.

Keywords: Facial Expression Recognition, Feature Extraction, Image Processing, Multilayer Perceptron Neural Network, Human-Computer Interaction.

1 Introduction

The aim of facial expression recognition is to enable machines to automatically estimate the emotional content of a human face. The facial expressions assist in various cognitive tasks and is well pointed that the expressions provide the most natural and powerful means for communicating human emotions, opinions and intentions. In an early work of Mehrabian [8] it has been indicated that during the face to face human communication only 7% of the information of a message is communicated by the linguistic (verbal) part of the message, i.e. spoken words, 38% by paralanguage (vocal part) and the 55% is communicated by the facial expressions. So, facial gestures constitute the most important communication part in face to face communication.

Analyzing the expressions of a human face and understanding their emotional state can find numerous applications to wide-ranging domains. The interaction between human and computer systems (HCI) would become much more natural and vivid if the computer applications could recognize and adapt to the emotional state of the human. Embodied conversational agents can greatly benefit from spotting and understanding the emotional states of the participants, achieving more realistic interactions at an emotional level. In intelligent tutoring systems, emotions and learning are inextricably

L. Iliadis et al. (Eds.): AIAI 2014, IFIP AICT 436, pp. 236–245, 2014.

bound together, and so recognizing the learner's emotional states could significantly improve the learning procedure delivered to him/her [1, 14]. Moreover, surveillance applications such as driver monitoring and elderly monitoring systems could benefit from a facial emotion recognition system, gaining the ability to deeper understand and adapt to the person's cognitive and emotional condition Also, facial emotion recognition could be applied to medical treatment to monitor patients and detect their status. However, the analysis of the human face characteristics and the recognition of its emotional state are considered to be very challenging and difficult tasks. The main difficulty comes from the non-uniform nature of the human face and various limitations such as lightening, shadows, facial pose and orientation conditions [6].

In order to classify facial human expression into the proper emotional categories, it is necessary to locate and extract the important facial features which contribute in identifying the expression's emotions. In this work, a facial emotion recognition system developed to determine the emotional state of human facial expressions is presented. Initially, the system analyzes the facial image, locates and measures distinctive facial deformations and characteristics such as the eyes, the eyebrows and the mouth. Each part of the face then is deeper analyzed and its features are extracted. The features are represented as information vectors. The classification of the feature vectors into the appropriate expression emotion is conducted by a multilayer neural network which is trained and used for the classification of the facial expression to the proper emotional category.

The rest of the paper is organized as follows. In section 2, basic topics on face image analysis and emotion recognition are described and also related work is presented. In Section 3, the methodology followed and the system developed is illustrated. In Section 4, the evaluation conducted and the experimental results gathered are presented. Finally, Section 5 concludes our paper and presents direction for future work.

2 Background and Related Work

2.1 Background Topics

In the field of facial emotion recognition two types of methods dominate: the holistic methods and the analytical or local-based methods [11]. The holistic methods try to model the human facial deformations globally, which encode the entire face as a whole. On the other hand, the analytical methods observe and measure local or distinctive human facial deformations such as eyes, eyebrows, nose, mouth etc. and their geometrical relationships in order to create descriptive and expressive models [2]. In the feature extraction process for expression analysis there are mainly two types of approaches which are the geometric feature based methods and the appearance based methods. The geometric facial features try to represent the geometrical characteristics of a facial part deformation such as the part's locations and model its shape. The appearance based methods utilize image filters such as Gabon wavelets to the whole face or on specific parts to extract feature vectors.

The way that emotions are represented is a basic aspect of an emotion recognition system. A very popular categorical model is the Ekman emotion model [4], which specifies six basic human emotions: anger, disgust, fear, happiness, sadness, surprise.

It has been used in several studies and systems that recognize emotional text and facial expressions related to these emotional states. Another popular model is the OCC (Ortony/Clore/Collins) model [10] which specifies 22 emotion categories based on emotional reactions to situations and is mainly designed to model human emotions in general. In this paper, an analytical approach is implemented for recognizing the basic emotions as defined by Ekman. More specifically, special areas of interest of the human face are analyzed and their geometrical characteristics such as locations, length, width and shape are extracted.

2.2 Related Work

Over the last decade there are a lot of efforts and works on facial image analysis and emotion recognition. A recent and detailed overview of approaches and methodologies can be found in [17, 18]. The work in [3], a facial expression classification method based on histogram sequence of feature vector is presented. It consists of four main tasks which are image pre-processing, mouth segmentation, feature extraction and classification which is based on histogram-based methods. The system is able to recognize five human expressions: happy, anger, sad, surprise and neutral based on the geometrical characteristics of the human mouth with an average recognition accuracy of 81.6%. In [9], authors recognize Ekman basic emotions sad, anger, disgust, fear, happy and surprise in facial expressions by utilizing Eigen spaces and using dimensionality reduction technique. The system developed achieved a recognition accuracy of 83%. The work presented in [16] recognizes facial emotions based on a novel approach using Canny, principal component analysis technique for local facial feature extraction and artificial neural network for the classification process. The average facial expression classification accuracy of the method is reported to be 85.7%. The authors in the work presented in [12], recognize four basic emotions of happiness, anger, surprise and sadness focusing in preprocessing techniques for feature extraction such as Gabor filters, linear discrimination analysis and Principal component analysis. They achieve in their experiments 93.8% average accuracy in images of the Jaffe face database with little noise and with particularly exaggerated expressions and an average accuracy of 79% in recognition on just smiling/non smiling expressions in the ORL database. In [15], a work that recognizes the seven emotions on Jaffe database using Fisher weight map is presented. Authors utilize image preprocessing techniques such as illumination correction and histogram equalization and the recognition rate of their approach is reported to be 69.7%.

3 Emotion Recognition System

In this section, we present the emotion recognition system developed and illustrate its functionality. The aim of the system is to endow software applications with the ability to recognize users' emotions in a similar way that the human brain does. The system takes as input images of human face and classifies them regarding the emotional state according to the basic emotions determined by Ekman. The system's steps for the automatic analysis of human facial expressions are presented in Figure 1.

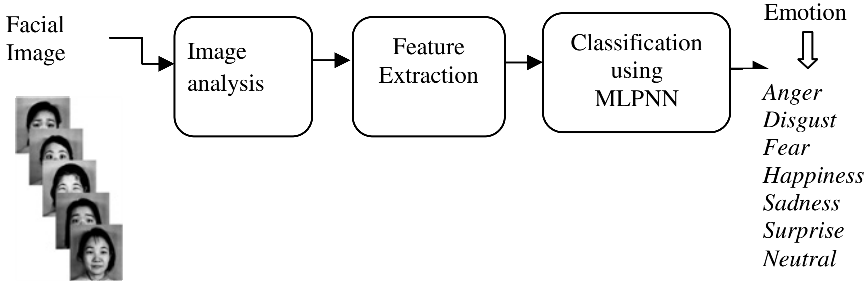

Fig. 1. The work flow chart of the system

Initially, the image analysis processes try to normalize the image and detect special regions of the face which contribute in identifying its emotional content. These regions are named Areas of Interests (AOIs) and are the region of the eyes, the area of the mouth and the eyebrows. The feature extraction process tries to select the proper facial features which describe the characteristic of the facial part and contribute in the classification stage. Since the system follows an analytical local-based approach, the feature extraction is implemented in the specific AOIs of the human face. Finally, the classification process, based on the feature extracted, specifies the appropriate emotional category among the set of different the emotional states defined by Ekman by utilizing a Multi-Layer Perceptron Neural Network (MLPNN) approach.

3.1 Image Analysis and AOIs Determination

Initially, the images of the human face are preprocessed using contrast limited histogram equalization. The histogram equalization is used to enhance the contrast in order to obtain a uniform histogram and also it re-organizes the image's intensity distributions and eliminates lighting effects [13]. A Sobel filter is then applied to emphasize edges and transition of the face and after its application, the image of the human face is represented as a matrix whose elements have only two possible values: zero and one. An element that has a value set to one represents an edge of a facial part in the image, such as an edge of an eye or the mouth. More specifically, the system analyzes a facial image as follows:

1. Apply histogram equalization to enhance the contrast of the image.
2. Apply a Sobel filter to emphasize edges and transition of the face.
3. Cut the face into eight horizontal zones.
4. Measure each zone's pixel density.
5. Select the zone with the highest pixel density. This zone contains the eyes and the eyebrows.
6. Locate the eyes and eyebrows.
7. Locate the mouth in the center of the zone.

A zone's pixel density is found by the average sum of consecutive rows of this matrix. So, in his way, pixel density of a zone shows the complexity of the face in each particular zone. The zone with the highest complexity contains the eyes and the eyebrows. The stages of the image analysis are presented in Figure 2.

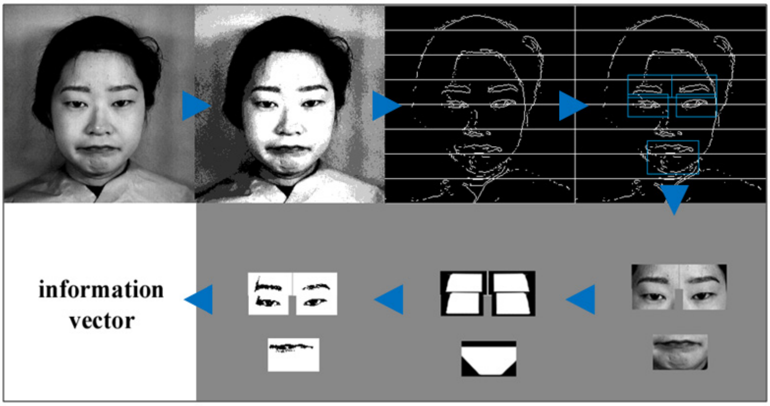

Fig. 2. The analysis stages of the facial image

3.2 AOI Analysis and Feature Extraction

The extraction of the proper facial feature is considered to be the most important step in facial expression recognition and is based on finding sets of features that convey meaningful information about the facial expression. Features extracted represent the geometrical characteristics of the facial part's deformation such as the part's height, width and model the element's shape. After the successful AOI isolation, the analysis of each AOI is performed in an effort to simplify the gathered information and extract the proper feature values. The analysis procedure implemented has seven steps:

1. Measure the average brightness of the face.
2. Set a brightness threshold.
3. Relate the brightness threshold to the average brightness.
4. Find the product of the brightness threshold and the average brightness.
5. Set a brightness filter.
6. Apply the brightness filter to the AOI.
7. Cut of the excess AOI noise using proper masks.

Application of an effective brightness filter has the effect to simplify an AOI in a way so all the data containing the emotional state of the face are left unaltered while all the irrelevant to the task data are discarded. The AOIs, after the filter application, are represented more smoothly and the information vectors can be easily extracted. To measure the average brightness, it is necessary to apply a facial mask to isolate the face area from the background image and the model's hair since these areas do not affect the brightness of the face. A series of 5 more sub masks must also be applied as a final step to reduce excess noise in the marginal areas of the AOI. The features extracted for each AOI are the following (see Figure 3):

Left eyebrow

- H1: The height of the far left part of the eyebrow.
- H2: The height of the part found on the 1/3 of the distance between the far left part and far right part of the eyebrow.
- H3: The height of the part found on the 2/3 of the distance between the far left and far right part of the eyebrow.
- H4: The height of the far left part of the eyebrow.
- L1: The length of the eyebrow.

Right eyebrow

- H5: The height of the far left last part of the eyebrow.
- H6: The height of the part found on the 1/3 of the distance between the far left and far right part of the eyebrow.
- H7: The height of the part found on the 2/3 of the distance between the far left and far right part of the eyebrow.
- H8: The height of the far left part of the eyebrow.
- L2: The length of the eyebrow.

Left eye

- H9: The height of the far left part of the eye.
- H10: The height of the part found on the 1/2 of the distance between the far left and far right part of the eye.
- H11: The height of the far right part of the eye.
- W1: The width of the eye.
- L3: The length of the eye.

Right eye

- H12: The height of the far right part of the eye.
- H13: The height of the part found on the 1/2 of the distance between the far left and far right part of the eye.
- H14: The height of the far left part of the eye.
- W2: The width of the eye.
- L4: The length of the eye.

Mouth

- H15: The height of the far left part of the mouth.
- H16: The height of the part found on the 1/2 of the distance between the far left and far right part of the mouth.
- H17: The height of the far left part of the mouth.
- W3: The width of the mouth.
- L5: The length of the mouth.

As mentioned above the recognition of the expressions' emotional content is determined based on the five AOIs. Each AOI is represented geometrically by five (5) values therefore the length of the information vector representing the facial expression is 25. The information vector contains all the information needed to classify a face into each of the seven basic facial emotional states.

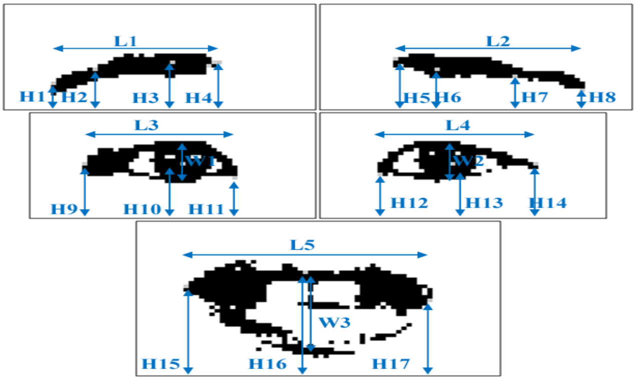

Fig. 3. Features extracted from each AOI

3.3 Neural Network Classification

The classification of the facial expression to the proper emotional category is made based on a Multi-Layer Perceptron Neural Network (MLPNN) which enables higher flexibility in training and also great adaptation to the problem of emotional classification [5]. The MLPNNs are feed-forward neural networks and can be used for the classification of multi-class, nonlinear problems. In our methodology, the MLPNN used for the classification of the facial expressions has three hidden layers, containing 13, 9 and 5 neurons respectively. Since there is no standard method to select the number of hidden layers and neurons due to the nature of the Neural Networks, the selection is often, as in this case, the architecture with the best performance. The input layer has 25 sensory neurons to match the length of the information vectors. The output layer has 3 neurons, that is, to classify the seven different classes. The output neurons combined produce a binary 3-digit sequence and each sequence is mapped to a class. The MLPNN network is trained using the back propagation supervised learning technique which provides a systematic way to update the synaptic weights of multi-layer perceptron networks.

4 Experimental Study

4.1 Data Collection

The methodology and the system developed were evaluated on facial images from the Japanese Female Facial Expression Database (JAFFE) [7]. Jaffe database contains 213 images of seven facial expressions angry, disgust, fear, happy, sadness, surprise and the neutral expression, posed by 10 different Japanese female models of young age. Each image has resolution of 256 x 256 pixels. The seven facial expressions (6 basic facial expressions and 1 neural expression) as posed of a JAFFE model are illustrated in Figure 4.

Fig. 4. The seven facial expressions as posed by a JAFEE model

A Face Matrix is represented as 256 x 256 pixel square matrix. The elements of the matrix represent the pixels of the Jaffe model's face. Each element of the face matrix ranges from 0 to 255. An element with a value between 0 and 255 represent a gray pixel of the image. An element with the value 0 represents a completely black pixel of the original image, while an element with the value 255 represents a completely white pixel, thus making the Face Matrix a Brightness Matrix.

4.2 System Evaluation

For the evaluation of the system's performance two datasets were created from the Jaffe database, the training dataset and the test dataset. The training dataset contains 140 facial images, 20 for each one of the seven facial emotions. The test dataset contains a total of 73 images, at least 9 images for each emotional state. The training dataset is inputted into the MLPNN, and the weights of the neurons are calculated in an optimal way. After the training phase is completed, the system's performance is evaluated on the test dataset. The results gathered are presented in Table 1.

Table 1. Performance results of the classification

Average accuracy	76,7 %
Precision	78,9%
F-measure	77,8%

The results show a good performance of the system. More specifically, from the corpus of 73 tested, it correctly identified the emotional state in 56 images. So, the general accuracy of the system is 76,7 % which indicates a good performance.

A noticeable point is that the system had a very good performance in identifying happiness emotion. This is due to the fact that in general and also in the database, this is a very strong emotion and is almost always expressed with vivid facial expression deformations. In contrast, the system lacks in recognizing the anger emotion. We believe that the main reason that the anger emotion is poorly classified is that the Jaffe models express this emotion in an inconsistent way regarding its valence. Specifically, in some cases, the models express anger emotion extremely, in an almost theatrical way while other models express the same emotion more natural in a smooth way.

Table 2. The confusion matrix

	Joy	Sadness	Surprise	Fear	Disgust	Anger	Neutral
Joy	12	0	0	0	1	0	0
Sadness	0	6	0	2	1	1	0
Surprise	0	0	8	0	0	0	1
Fear	0	0	0	9	0	0	2
Disgust	0	1	0	0	8	0	0
Anger	0	1	0	0	3	6	1
Neutral	0	0	0	1	2	0	7

5 Conclusion and Future Work

In this paper, an automatic facial expression recognition system is presented. The aim of the system developed is to analyse and recognize the emotional content of human facial expressions. The system initially analyzes the facial image, locates and measures distinctive human facial deformations such as the eyes, the eyebrows and area of the mouth and extracts the proper features. The features extracted try το model the geometrical characteristics of each area part of the human face. Finally, for the categorization, a multilayer perceptron neural network is trained and used. The experimental study conducted indicates quite promising performance.

However, there are some points that the system could be improved. Initially, evaluating the system on different facial expression databases such as the ORL and the Cohn-Kanade facial expression databases will give us a better insight of or methodology's performance. Moreover, currently the face analysis and the determination of the AOIs work on still frontal images of human face and so, a challenging extension will be to detect and analyse facial expressions from different poses. Furthermore, the system's classification approach is based on a multilayer perceptron neural network. Extending the classification method by using a neuro-fuzzy approach could improve the system's accuracy in recognising the emotional state. Exploring this direction is a key aspect of our future work.

References

1. Akputu, K.O., Seng, K.P., Lee, Y.L.: Facial Emotion Recognition for Intelligent Tutoring Environment. In: 2nd International Conference on Machine Learning and Computer Science (IMLCS 2013), pp. 9–13 (2013)
2. Arca, S., Campadelli, P., Lanzarotti, R.: An automatic feature-based face recognition system. In: Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004) (2004)
3. Aung, D.M., Aye, N.: A Facial Expression Classification using Histogram Based Method. In: International conference on Signal Processing Systems (2012)
4. Ekman, P.: Basic emotions. Handbook of Cognition and Emotion, pp. 45–60 (1999)

5. Filko, D., Goran, M.: Emotion recognition system by a neural network based facial expression analysis. Autom. –J. Control Meas. Electron. Comput. Commun. 54, 263–272 (2013)
6. Koutlas, A., Fotiadis, D.I.: An automatic region based methodology for facial expression recognition. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2008, pp. 662–666 (2008)
7. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205. IEEE Computer Society, Los Alamitos (1998)
8. Mehrabian, A.: Communication without words. Psychology Today 2(4), 53–56 (1968)
9. Murthy, G.R.S., Jadon, R.S.: Recognizing facial expressions using eigenspaces. In: International Conference on Conference on Computational Intelligence and Multimedia Applications, vol. 3, pp. 201–207. IEEE (December 2007)
10. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
11. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1424–1445 (2000)
12. Přinosil, J., Smékal, Z., Esposito, A.: Combining features for recognizing emotional facial expressions in static images. In: Esposito, A., Bourbakis, N.G., Avouris, N., Hatzilygeroudis, I. (eds.) HH and HM Interaction. LNCS (LNAI), vol. 5042, pp. 56–69. Springer, Heidelberg (2008)
13. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing using MATLAB. Prentice Hall (2003)
14. Shen, L., Wang, M., Shen, R.: Affective e - learning: Using “emotional” data to improve learning in pervasive learning environment. Educational Technology & Society 12(2), 176–189 (2009)
15. Shinohara, Y., Otsu, N.: Facial expression recognition using fisher weight maps. In: Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 499–504. IEEE (May 2004)
16. Thai, L.H., Nguyen, N.D.T., Hai, T.S.: A facial expression classification system integrating canny, principal component analysis and artificial neural network. arXiv preprint arXiv:1111.4052 (2011)
17. Tian, Y., Kanade, T., Cohn, J.F.: Facial expression recognition. In: Handbook of Face Recognition, pp. 487–519. Springer, London (2011)
18. Verma, A., Sharma, L.K.: A Comprehensive Survey on Human Facial Expression Detection. International Journal of Image Processing (IJIP) 7(2), 171 (2013)