Recognize Emotions from Facial Expressions Using a SVM and Neural Network Schema

Isidoros Perikos(✉), Epaminondas Ziakopoulos, and Ioannis Hatzilygeroudis

School of Engineering, Department of Computer Engineering & Informatics, University of Patras, 26504, Patras, Greece
{perikos,ziakopoulo,ihatz}@ceid.upatras.gr

Abstract. Emotions are important and meaningful aspects of human behaviour. Analyzing facial expressions and recognizing their emotional state is a challenging task with wide ranging applications. In this paper, we present an emotion recognition system, which recognizes basic emotional states in facial expressions. Initially, it detects human faces in images using the Viola-Jones algorithm. Then, it locates and measures characteristics of specific regions of the facial expression such as eyes, eyebrows and mouth, and extracts proper geometrical characteristics form each region. These extracted features represent the facial expression and based on them a classification schema, which consists of a Support Vector Machine (SVM) and a Multilayer Perceptron Neural Network (MLPNN), recognizes each expression's emotional content. The classification schema initially recognizes whether the expression is emotional and then recognizes the specific emotions conveyed. The evaluation conducted on JAFFE and Kohn Kanade databases, revealed very encouraging results.

Keywords: Emotion recognition · Facial gestures · Human-computer interaction · Support vector machines · Multilayer Perceptron Neural Network

1 Introduction

Facial expressions form a universal language of emotions, which can instantly express a wide range of human emotional states and feelings. The facial expressions assist in various cognitive tasks; so, reading and interpreting the emotional content of human expressions is essential to deeper understand the human condition. In an early work on human facial emotions [10] it has been indicated that during a face-to-face human communication only 7% of the information of a message is communicated by the linguistic part of the message, such as spoken words, 38% is communicated by paralanguage (vocal part) and 55% is communicated by the facial expressions. So, the facial expressions constitute the most important communication medium in face-to-face interaction.

Giving to computer applications the ability to recognize the emotional state of humans from their facial expressions is a very important and challenging task with wide ranging applications. Indeed, automated systems that can determine emotions of a human based on his/her facial expressions, can improve the human computer

L. Iliadis and C. Jayne (Eds.): EANN 2015, CCIS 517, pp. 265–274, 2015.
DOI: 10.1007/978-3-319-23983-5_25

interaction and give computer systems the opportunity to customize and adapt its response. In intelligent tutoring systems, emotions and learning are inextricably bound together; so, recognizing the learner's emotional states could significantly improve the efficiency of the learning procedures 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 recognition of the facial emotional state is considered to be very challenging task, due to the non-uniform nature of the human face and various conditions such as, lightening, shadows, facial pose and orientation [7].

In this work, a facial emotion recognition system developed to determine the emotional state of human facial expressions is presented. The system analyzes facial expressions and recognizes the six basic emotions as defined by Ekman [4], following an analytical, local-based approach. Initially, given a new image, the system detects human faces in it using the Viola-Jones algorithm [18]. Then, it locates and measures facial deformations of specific regions such as eyes, eyebrows and mouth and extracts geometrical characteristics such as locations, length, width and shape. These extracted features represent the deformations of the facial expression and based on them a classification schema, which consists of a Support Vector Machine (SVM) and a Multilayer Perceptron Neural Network (MLPNN) classifier, recognizes each expression's emotional content. Initially, the SVM classifier characterizes the facial expression as emotional or neutral and then the MLPNN recognizes and classifies the emotional facial expression into the proper emotion category. The evaluation study was conducted on JAFFE and Kohn Kanade databases and revealed very encouraging results regarding the performance of the system in recognizing emotional expressions and specifying their emotional content on Ekman's scale.

The structure of the reminder of the paper is as follows. In section 2, related work is presented. In Section 3, the methodology and the system developed are illustrated. In Section 4, the evaluation conducted and the experimental results gathered are presented. Finally, Section 5 concludes the paper and presents direction for future work.

2 Related Work

In the literature there are a lot of research efforts on facial image analysis and emotion recognition [3][17]. Although a human can detect and interpret faces and facial expressions naturally with little or no effort, accurate facial expression recognition by machines is still a challenge. Authors, in the work presented in [2], present a method for emotional classification of facial expressions, which is based on histogram sequence of feature vector. The system is able to recognize five human expression categories: happy, anger, sad, surprise and neutral, based on the geometrical characteristics of the human mouth with an average recognition accuracy of 81.6%. The work presented in [15] recognizes facial emotions based on a novel approach using Canny, principal component analysis (PCA) technique for local facial feature extraction and an 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 [13], recognize four basic emotions: happiness, anger, surprise and sadness, focusing in preprocessing techniques for feature extraction such as, Gabor filters, linear discrimination analysis and PCA. They achieve in their experiments 93.8% average accuracy for images of the Jaffe face database with little noise and with particularly exaggerated expressions and an average accuracy of 79% in recognition of just smiling/non smiling expressions in the ORL database. In [11], authors use a SVM classifier to recognize the 6 basis Ekman emotions in Cohn-Kanade database. In the study, authors extract 22 features form facial still images and report an average accuracy of 87.9%. In [8], authors recognize the six basic Ekman's emotion in elders' facial expressions using a SVM classifier. Features are extracted for expressions based on the active shape model, which fits in the shape parameters, using techniques such as gradient descent. The SVM is trained on a dataset such as, JAFFE in Cohn-Kanade and MII, and authors report promising results. In [19], authors developed a system for the emotion classification through lower facial expressions using the adaptive SVM. The system extracts eight feature points from the mouth, chin and nose and feed them into the A-SVMs classifier to categorize expression. The system was tested on Jaffe database and reports an average accuracy of 74.5%.

3 Emotion Recognition System

The system, given a new image, analyzes it and tries to detect human faces in it using the Viola-Jones method [18]. Then, it analyzes the facial expression, extracts appropriate features and classifies it to the proper emotion category, as defined by Ekman. The steps are depicted in Figure 1.

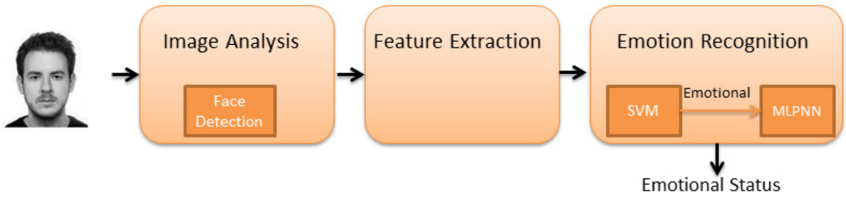

Fig. 1. The work flow chart of the system

When a face is detected in the image, the area of the face is located and is further analyzed. Initially, the facial area is normalized and the contrast is enhanced and after that, specific regions which contribute in recognizing the emotional content of the facial expression are specified. The specific regions, from which the proper features are extracted, 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 analyzes the AOIs and tries to extract proper facial features, which describe and model the characteristics of the facial region. The system follows an analytical local-based approach and so the feature extraction is implemented in the specific AOIs of the face. Finally, the emotion recognition is conducted by the classification schema, which specifies the appropriate emotional content of the expression on Ekman's scale. The classification schema consists of a Support Vector Machine (SVM) and a Multi-Layer Perceptron Neural Network (MLPNN) classifier, where the SVM specifies whether a facial expression is emotional or not and the MLPNN recognize the emotional content of the expressions recognized as emotional by the SVM.

3.1 Image Analysis and Face Detection

The system utilizes the Viola - Jones algorithm to locate human faces in an image. In general, the height and width of the face-box output by the Viola and Jones face detector are determined accurately and can vary by approximately 5% [16]. In case that more than one faces are detected, each face is analyzed separately. A human face is further analyzed in order to locate specific AOIs. The first step, aiming to enhance the contrast of the face, is to apply histogram equalization and then a sober filter to emphasize the edges of the face. Then, the exact location of the eyes is specified. The locations of the rest of the AOIs are detected based on the relative position of the eyes: the eyebrows are located right above the eyes and the mouth is located in the horizontal center of the face under the eyes. More specifically, the image is analyzed as follows:

1. Detect the face in the image using the Viola - Jones algorithm.
2. Apply histogram equalization to enhance contrast of the upper-half of the face, which contains the eyes.
3. Apply a Sobel filter to emphasize edges and transition of the face.
4. Cut the face into four horizontal zones.
5. Measure each zone's pixel density by averaging the number of the white (transitional) pixels in each individual zone.
6. Select the zone with the highest pixel density. This zone contains the eyes.
7. Locate eyebrows and mouth using facial features mask (new figure). The eyebrows are located right above the eyes and the mouth is located under the eyes.

The histogram equalization is very useful and is used to enhance the contrast of the image in order to obtain a uniform histogram and re-organizes the image's intensity distributions. After that, a Sobel filter is 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 of the mouth etc. The pixel density of a zone is calculated to be the average sum of consecutive rows of the matrix of the zone. In this approach, the pixel density of a zone represents the complexity of the face in each particular zone. The zone with the highest complexity contains the eyes and the eyebrows. In Figure 2, the stages of the image analysis and the results of each process step are illustrated in an example facial image.

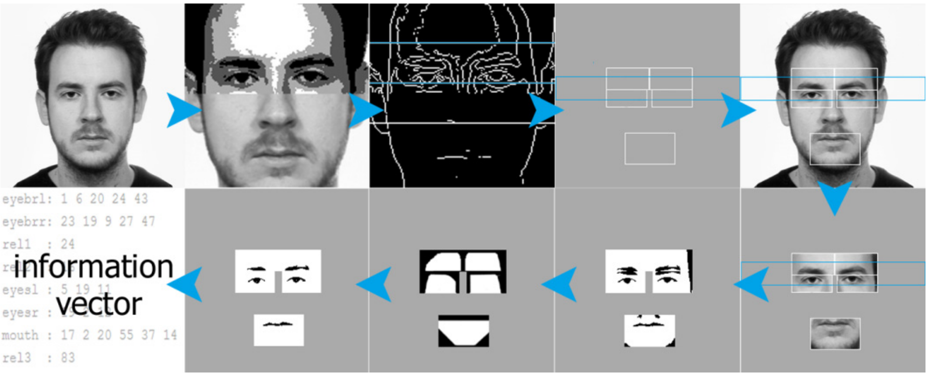

Fig. 2. The stages of the analysis of a facial image.

3.2 AOI Analysis and Feature Extraction

After the AOIs of the human face have been located, each AOI is further analyzed in order to extract proper features that describe its characteristics and convey meaningful information about the AOI deformation. The features extracted represent the geometrical characteristics of the AOI's deformation such as the facial part's height and width, and also try to 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 steps of the AOIs analysis are the following:

1. Measure the average brightness of the face.
2. Set a brightness threshold. All the elements of the matrix with higher values than the threshold maximize their values and all the elements of the matrix with lower values than the threshold minimize their values
3. Relate the brightness threshold to the average brightness. Brighter faces require higher thresholds than darker faces, in order to minimize useful data loss from the application of the filter.
4. Set the brightness filter.
5. Apply the brightness filter to the AOI.
6. Cut of the excess AOI noise using proper individual masks.

Initially, a brightness filter is applied to simplify an AOI in order to remove unnecessary data. AOIs, after the filter application, are represented more smoothly and the features can be easily extracted. In order to measure the average brightness, a facial mask is applied to isolate the face area from the background image and also from the model's hair since these areas do not affect the brightness of the face. In Figure 3, the features extracted from each AOI in the example facial image presented above are illustrated.

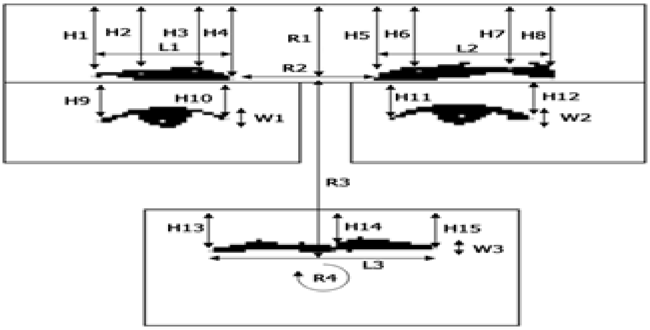

Fig. 3. Features extracted from each AOI

Left eyebrow

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

Right eyebrow

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

Left eye

- H9: The height of the far left part.
- H10: The height of the far right part.
- W1: The width of the eye.

Right eye

- H11: The height of the far left part.
- H12: The height of the far right part.
- W2: The width of the eye.

Mouth

- H13: The height of the far left part of the mouth.
- H14: The height of the part found on the 1/2 of the distance between the far left and far right part of the mouth.
- H15: The height of the far left part of the mouth.
- W3: The width of the mouth.
- L3: The length of the mouth.

Relative Values

- R1: The average height of the eyebrows.
- R2: The horizontal distance between the eyebrows.
- R3: The vertical distance between the eyebrows and the bottom of the mouth.
- R4: The ratio of the mouth length to the mouth width.

So, a facial expression is represented by 25 features. These features model the facial expression and contain the necessary data to assist the classifiers to classify an expression into the proper category.

3.3 Classification Schema

The classification schema consists of two classifiers, a SVM and a MLPNN. SVMs are very powerful binary classifiers. In general, a SVM classifier constructs a hyperplane or a set of hyperplanes in a high- or infinite-dimensional space, which can be used for the classification procedure. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the SVM classifier. For the purposes of our work, the SVM is used to separate the facial expressions that convey emotion from the neutral ones. The second classification method is based on a MLPNN classifier which enables higher flexibility in training and also great adaptation to the problem of emotional classification [5]. In general, MLPNNs are feed-forward neural networks and can be used for solving of various multi-class, nonlinear classification problems. In our methodology, the MLPNN classifier used for the emotional classification of the facial expressions has three hidden layers, containing 14, 14 and 8 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 of the MLPNN has 25 sensory neurons, in order to match the length of the information vectors. 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. Both the MLPNN and the SVM classifiers were implemented in MATLAB toolkit.

4 Experimental Study

4.1 Data Collection

The methodology and the system developed were extensively evaluated on facial images from two popular and widely used databases, the Japanese Female Facial Expression Database (JAFFE) [9] and the Cohn-Kanade database [6]. The Jaffe database consists of 211 facial pictures of 10 different posers. Approximately the two thirds (140 images) of the images of the database were selected to be part of the training dataset and the remaining 71 images were part of the test dataset. The Cohn-Kanade is a popular database which includes 486 sequences from 97 posers. For the needs and purposes of this study, the total number of the selected images was 209, where 140 images were part of the training dataset and 69 images were part of the test dataset respectively.

4.2 System Evaluation

Initially, an evaluation of the system's performance in characterizing a facial expression either as emotional or as neutral was conducted. The performance of the mechanism is presented in Table 1.

Table 1. SVM Performance Results

Metric	Jaffe	Kohn Kanade	Total
Accuracy	92.3	100	96.4
Precision	98.3	100	99.2
Sensitivity	93.4	100	96.9
Specificity	90	100	95

The results indicate that the system has a very good performance in determining whether a facial expression conveys emotions or not. Indeed, the SVM achieves excellent performance in the Kohn Kanade database, mainly due to the fact that Kohn Kanade models present neutral expressions in a very consisted and inactive way. Also, high sensitivity indicates that the mechanism can accurately identify emotional gestures that indeed are emotional.

The second part of the evaluation study assessed the performance of the MLPNN mechanism of the system in recognizing the emotional content of an emotional facial expression. The facial expressions that were emotional and correctly recognized as emotional by the SVM mechanism, were used for this part of the experiment. Given that we have a multiple class output, we use the following metrics: average accuracy, precision and f-measure. The performance of the system is presented in Table 2.

Table 2. Performance Results of the MLPNN

Metric	Jaffe	Kohn Kanade	Total
Average Accuracy	89,7 %	81,4%	85,5%
Precision	87,3%	82,8%	86,7%
F-measure	88,9%	81.4%	86,7%

The MLPNN demonstrates quite satisfactory performance in recognizing the emotional status of emotional expressions. The results are better in the JAFFE database and we believe that the reason is the strength that the JAFFE models express the emotions. Indeed, in some cases the models of the JAFFE database express the different emotions very vividly and with strong facial deformations.

The third part of the study evaluated the overall performance of the system developed. The overall performance of the system and the confusion matrix of the performance on both databases are presented in Table 3 and Table 4 respectively.

The overall evaluation results indicate that the system is performing very well in both databases. A noticeable point concerns the high performance of the system in recognizing facial gestures that express happiness. This is mainly due to the fact that joy is a very strong emotion that in most cases is expressed by vivid facial deformations.

Table 3. The Performance Results of the System

Metric	Jaffe	Kohn Kanade	Total
Average Accuracy	85.9 %	84,1%	85%
Precision	87,5%	85,6%	86%
F-measure	87.4%	85.4%	86.1%

Table 4. The Combined Confusion Matrix of System's Performance.

	Joy	Sadness	Surprise	Fear	Disgust	Anger	Neutral
Joy	22	0	0	0	0	1	0
Sadness	0	12	0	6	0	1	0
Surprise	0	0	14	3	0	1	1
Fear	0	3	0	16	0	0	1
Disgust	0	3	0	2	13	1	0
Anger	0	1	0	1	0	17	1
Neutral	0	0	2	2	1	0	15

5 Conclusion and Future Work

In this paper, we present an emotion recognition system, which recognizes the six basic emotions defined by Ekman in facial expressions. Initially, it detects human faces in images using the Viola-Jones algorithm. Then it locates and measures characteristics of specific regions and extracts proper features and geometrical aspects form each region. These extracted features represent the facial expression and based on them a classification schema recognizes each expression's emotional content. The classification schema initially recognizes whether the expression is emotional, based on a SVM classifier, and then recognizes the specific emotions conveyed by the expression, based on a MLPNN. The evaluation conducted on JAFFE and Kohn Kanade databases revealed very encouraging results.

As a future work, we plan to conduct a bigger scale evaluation and also examine the system's performance in additional databases. Furthermore, an extension could be made in order to assist the system to analyse expressions from different poses and face orientations. Finally, the classification schema could be extended and the exploration of a neuro-fuzzy approach could enhance the system's classification performance. Exploring this direction is a main 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. Aung, D.M., Aye, N.A.: Facial expression classification using histogram based method. In: International Conference on Signal Processing Systems (2012)
3. Bettadapura, V.: Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722 (2012)
4. Ekman, P.: Basic emotions. In: 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. Kanade, T., Cohn, J.F., Tian, Y: Comprehensive database for facial expression analysis. In: Proceedings of the Automatic Face and Gesture Recognition (2000)
7. 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)
8. Lozano-Monasor, E., López, M.T., Fernández-Caballero, A., Vigo-Bustos, F.: Facial expression recognition from webcam based on active shape models and support vector machines. In: Pecchia, L., Chen, L.L., Nugent, C., Bravo, J. (eds.) IWAAL 2014. LNCS, vol. 8868, pp. 147–154. Springer, Heidelberg (2014)
9. 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)
10. Mehrabian, A.: Communication without words. Psychology Today **2**(4), 53–56 (1968)
11. Michel, P., El Kaliouby, R.: Facial expression recognition using support vector machines. In: The 10th International Conference on Human-Computer Interaction, Greece (2005)
12. Perikos, I., Ziakopoulos, E., Hatzilygeroudis, I.: Recognizing emotions from facial expressions using neural network. In: Iliadis, L. (ed.) AIAI 2014. IFIP AICT, vol. 436, pp. 236–245. Springer, Heidelberg (2014)
13. 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)
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. 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)
16. Valstar, M.F., Mehu, M., Jiang, B., Pantic, M., Scherer, K.: Meta-analysis of the first facial expression recognition challenge. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **42**(4), 966–979 (2012)
17. Verma, A., Sharma, L.K.: A Comprehensive Survey on Human Facial Expression Detection. International Journal of Image Processing (IJIP) **7**(2), 171 (2013)
18. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision **57**(2), 137–154 (2004)
19. Visutsak, P.: Emotion Classification through Lower Facial Expressions using Adaptive Support Vector Machines. JMMT: Journal of Man, Machine and Technology **2**(1), 12–20 (2013)