Facial Expression Recognition Using Interpolation Features

Jesús García-Ramírez, Ivan Olmos-Pineda, J. Arturo Olvera-López, and
Manuel Martín-Ortíz

Benemérita Universidad Autónoma de Puebla, Faculty of Computer Science, Puebla, México
gr_jesus@outlook.com, {iolmos,aolvera,mmartin}@cs.buap.mx

Abstract. In this work, a methodology for classifying emotions (such as happiness, anger and surprise) based on face images is proposed. This methodology consist of three stages: in the pre-processing stage, edge detectors and threshold algorithms are used in order to find edge information about ROIs; in the second stage (feature extraction) numeric information of pre-processing images is extracted via interpolation methods; finally, in the classification stage supervised learners such as Neural Networks, Support Vector Machines and Naïve Bayes are tested. According to the results, our approach has acceptable accuracy in order to recognize emotions.

Keywords: Facial expression recognition, Face image pre-processing, Image segmentation, Supervised classification.

1 Introduction

Nowadays, data processing is an active research area in computer science, where subfields such as image processing is the baseline of computer vision. An application of computer vision is Facial Expression Recognition (FER) for detecting mental disorders, emotions, whether somebody is lying, among others [1]. A FER system detects six universal emotions (anger, happiness, disgust, surprise, sadness and fear) introduced by Paul Ekman in 1994 [2].

A FER system includes different stages (Fig. 1), starting with the pre-processing one where the input images (commonly in RGB color model) are processed applying different edge detectors and image filters in order to find descriptive data from the Regions of Interest (ROIs) in face images (eyebrows, eyes, mouth and nose). The next stage is the feature extraction, where the obtained information from pre-processing stage is used to get numeric information about ROIs. Finally, it is possible to use machine learning algorithms with the aim to classify new face images [3].

Several FER approaches have been reported in the literature, some of them based on Local Binary Pattern (LBP), using a histogram of the intensity values around a pixel to classify emotions via Support Vector Machines (SVM) [4]. Other approach is based on edge detection and parallel computing with CUDA (Compute Unified Device Architecture), in this work a CUDA implementation is used to detect borders for

Jesús García-Ramírez, Ivan Olmos-Pineda, J. Arturo Olvera-López, Manuel Martín-Ortíz

segmenting ROIs, then different points from each ROI were used for classification via SVM [5]. Sebe et al. [7] report a comparison among classifiers to determine emotions, particularly: Bayesian networks, nearest neighbors, decision threes and Naïve Bayes; according to the reported experiments, the best accuracy was obtained by k-nearest neighbors.
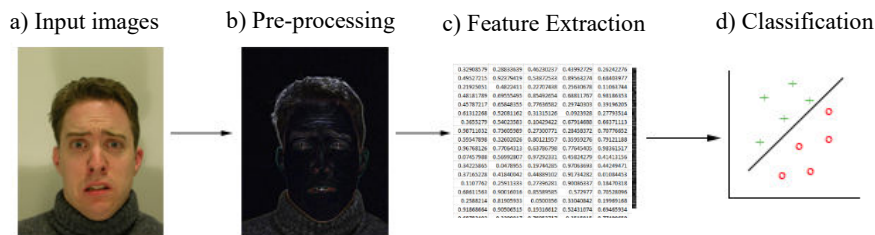
a) Input images   b) Pre-processing   c) Feature Extraction   d) Classification

Fig. 1. FER process a) input images in RGB color model or gray scale (face image extracted from [8]); b) image pre-processing applying edge detectors and filters in order to get descriptive information from the ROIs; c) feature extraction getting numeric data from pre-processing stage; d) classification process using data extracted on c) to get a model found by a classifier

In recent years, deep learning approaches have been proposed in order to get a higher accuracy in classification stage. A method based on Convolutional Neural Networks (CNN) is proposed by Wen G. et al. [9], they use four datasets resizing each image to 48 x 48 pixels, then these images are used as input of the CNN. Other approach uses ROIs such as nose, eyes, mouth or landmarks from these ROIs as input of a CNN, finally a good performance in classification stage is reported [10, 6]. Deep belief networks have been also applied to facial expression recognition where face images are split and those are the input of the learners (SVM and Adaboost classifiers) [11].

In this paper we propose a method for detecting three of the six universal emotions (anger, happiness and surprise), based on eyebrows and mouth features extracted via interpolation polynomials as descriptive information of ROIs. Finally, classification algorithms such as Neural Networks (NN) with feed forward architecture, SVM and Naïve Bayes were tested in order to compare their performance.

This paper is structured as follows: section 2 introduces the proposed methodology and each stage of the process is described; in section 3 the results of the classification process are shown; finally in section 4 the conclusions and future work are addressed.

2   Facial Expression Recognition Methodology

The proposed methodology is shown in Fig. 1, we assume that inputs are images in RGB color model. In feature extraction stage, eyebrows and mouth are processed because these ROIs provide information about facial expressions. Points and coefficients found with interpolated polynomials are used as features of the ROIs. Finally, in the classification process SVM, NN and Naïve Bayes are trained in order to test

their performance for classifying emotions. In the following sections each stage of our methodology is described.

2.1 Pre-processing

The pre-processing stage is based on edge detectors (such as gradient in different directions) and threshold operators (like Otsu [12]) in order to obtain intensity transition information from ROIs. In addition, unsupervised learners using RGB color model are tested for segmenting ROIs.

The RGB color model can be seen as a three-dimensional space, each one of the axis represents a color channel in the model (Red, Green and Blue), the intensity values of each pixel is denoted as a vector of three values, then a clustering algorithm can be used to segment images; face images can be segmented via clustering algorithms because of intensity values of the ROIs (in some cases) could be different respect to other regions. For the experimentation in this stage, different clustering algorithms are used (DBScan [13], OPTICS [14] and CLARANS [15]). We use as input, images from the MMI facial expression database [8]. As first step we find the face region through Viola & Jones algorithm [16], then the images are resized to 100 x 100 pixels. Before applying the clustering algorithms, a pre-processing stage by smoothing filters (median and mean) and automatic contrast adjustment considering frequency values higher than the 5% of the number of pixels in the image is applied.

In clustering process we apply three algorithms: DBscan, an algorithm based on detecting dense regions taking into account a minimum number of points to consider a region as dense (*min*) and a value (*e*) that is used to define the radius for the region; OPTICS works in a similar way as DBscan does, but OPTICS is order-based to analyze the regions. For both clustering algorithms in our experiments, *min* and *e* values were fixed to 3 and 5 respectively. Other applied clustering algorithm is CLARANS, this algorithm consist of analyzing sets of centroids, in this algorithm the number of clusters is five, the neighbors analyzed (sets that vary in at least one element) is three and the local minimum (used as stop condition) is one hundred. In Fig. 2 it can be seen the clustering results of each one of the algorithms applied over some images taken from the MMI database without applying a pre-processing, on the other hand, Fig. 3 shows the same images with a pre-processing stage consisting of apply smoothing filters (mean and median) and automatic contrast adjustment, due this contrast adjust noise is highlighted in the histogram. For both images (Fig. 2 and Fig. 3), the first row shows the input images, in the second row it is shown the output of DBscan algorithm. The output of CLARANS can be seen in the third row. Finally, in the fourth row the output of OPTICS algorithm is presented.

In general, clustering algorithms have not a good performance in order to get descriptive information about the ROIs, for this reason a different approach must be proposed to get better information about ROIs in face images.

In order to get a better performance than the experiments with the clustering algorithms for extracting ROIs, edge detectors and thresholding approaches were used to get information about eyebrows and mouth regions, for mouth region gradient edge detection in two directions (vertical and horizontal) is applied, finding the distance

between these values, then the image is denoised using its histogram and only the frequency values higher than the mean are selected, then a binarization process (*x'=0 if x<threshold, otherwise x'=255*) with *threshold=127* is applied. Finally, the black regions in the binary images are detected via metrics as density (black pixels respect to a rectangular area) and the rectangular area of the cluster related to original image size are considered to find the ROI in the image.

For eyebrow segmentation a different approach need to be improved, in this process a closing morphological operation is applied to smooth the border and remove thin holes in eyebrow region, then an operator is used to remove the shadow between eye and eyebrow, this consist in a threshold value, just considering higher values than this threshold over which sine function is applied, after that Otsu thresholding is carried out. Finally, the same process implemented for detecting regions in mouth is used for the eyebrow case.

In Fig. 4 the segmentation results of the approach described in the above paragraphs are shown, for each image the result for mouth and eyebrow regions is presented with a binary image. The segmented ROIs will be used in the next stage in order to extract numeric values for training a classifier.

2.2   Feature Extraction

In this stage two sets of features are extracted, for mouth region two quadratic polynomials are interpolated, one for the upper part and one for the lower part of the mouth. On the other hand, for eyebrow region only one cubic polynomial is found. The polynomials are found with the Newton divided differences method [17].

In Fig. 5 the process to obtain the features is depicted. The points are transferred to the origin with a vector and the coefficients for each region are found (CF, Coefficient Features), then other set of features are extracted from the polynomial, ten points with each polynomial are found (PF, Point Features). This two sets of features are used in the next stage in order to train a classifier.

2.3   Classification

In the classification process a supervised learner is used to get a model, in this stage we use three different learners for numeric attributes: Naïve Bayes, SVM and NN, which were trained using the features described in the section 2.2.

3   Experimental Results

In this section, the results of the classification process are presented, we use Naïve Bayes, SVM and NN learners and three of the six universal emotions are classified. The training set contains 42 instances for happiness, 17 instances for anger and 20 instances for surprise. For testing the accuracy results, 5-fold-cross-validation is applied in these experiments, all the attributes are normalized via z-score method.
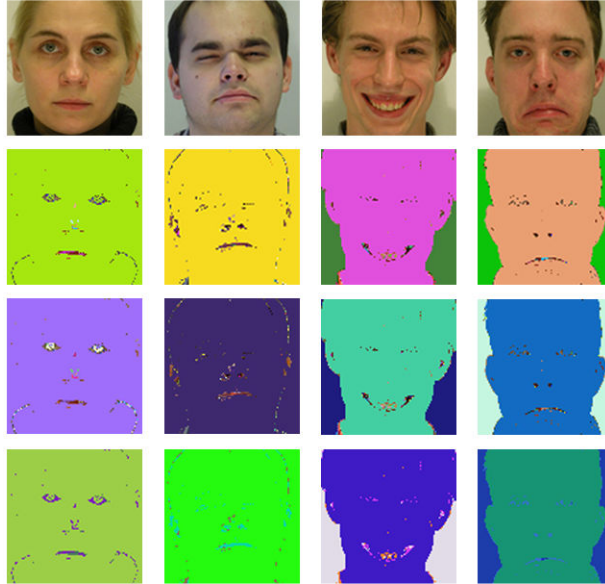
Fig. 2. Obtained results applying clustering algorithms without a pre-processing stage: first row shows input images; second row output of DBscan; in the third row CLARANS result can be seen; OPTICS result images are shown in fourth row

Fig. 3. Results of clustering algorithms applied over pre-processed images. The first row shows the input images; the other three rows show the results of clustering algorithms: second row OPTICS; third row CLARANS; fourth row BDscan
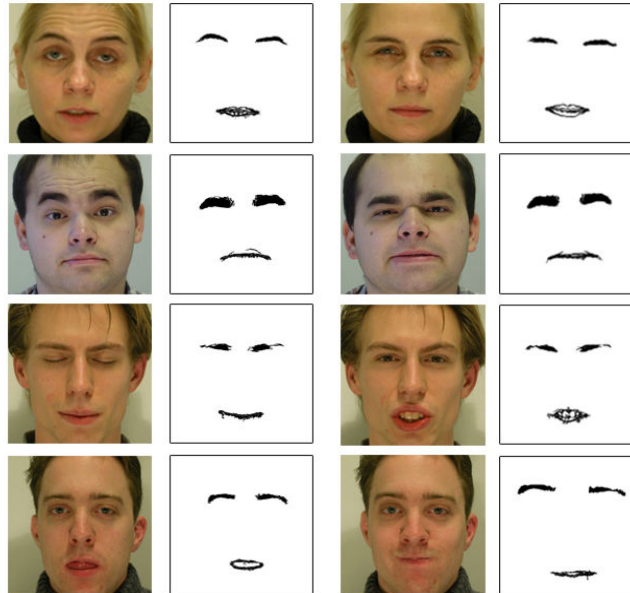
Jesús García-Ramírez, Ivan Olmos-Pineda, J. Arturo Olvera-López, Manuel Martín-Ortíz

Fig. 4. ROIs (eyebrow and mouth) segmentation results of some images taken from MMI database
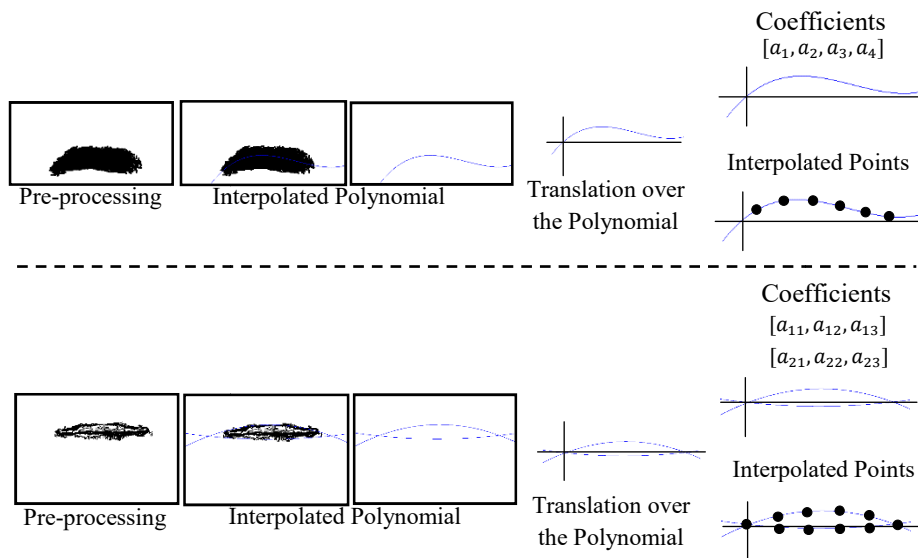

Fig. 5. Feature extraction for eyebrows and mouth regions based on interpolation polynomials

In order to get a better training for the Neural Network (feed forward architecture), different parameter values such as momentum, learning rate and neurons in the hidden layer were tested.

Our first experiment with the NN classifier uses as input the PF set explained in section 2.2. In Table 1 some of the tested parameters such as epochs, number of neurons in the hidden layer (in the experiments one hidden layer is used), momentum and learning rate are shown with their accuracy.

Table 1. Best experimented parameters for the NN classifier with PF

Epochs	Neurons in the hidden layer	Momentum	Learning Rate	Accuracy
500	**39**	**0.3**	**0.06**	**92.4%**
500	**44**	**0.4**	**0.06**	**92.4%**
500	41	0.1	0.27	91.8%
1000	45	0.16	0.58	92.3%
1500	45	0.11	0.6	91.9%

It can be seen that the first two rows in Table 1 the accuracy is 92.4%, there is a difference in the confusion of the emotions, the confusion matrix for the two set of parameters can be seen in Fig. 6, where H=Happiness, A=Anger and S=Surprise. The confusion matrices in Fig. 6 have the same accuracy, in Fig 6a some S cases are classified as A, meanwhile in Fig. 6b some cases S are confused with H and A.
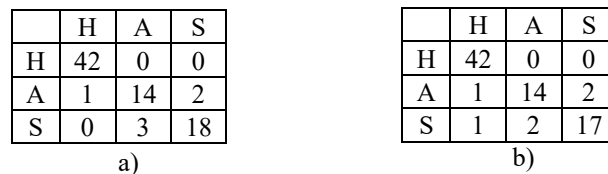
	H	A	S
H	42	0	0
A	1	14	2
S	0	3	18

a)

	H	A	S
H	42	0	0
A	1	14	2
S	1	2	17

b)

Fig. 6. Confusion matrices for experimented parameters with NN classifier and PF: a) parameters for the first row in Table 1; b) parameters for the second row in Table 1

The confusion matrices shown in Fig. 7 were obtained by classifying emotions using our best Neural Network configuration; Fig. 7a and 7b shows the confusion matrices for two experiments where NN is trained with CF, and the two best parameters described in Table 2, rows 3 and 1 respectively. Based on these results, it is possible to conclude that the best performance is obtained using PF as input features.

Table 2. Best experimented parameters for the NN classier with CF

Epochs	Neurons in the hidden layer	Momentum	Learning Rate	Accuracy
500	41	0.01	0.11	83.54%
500	36	0.13	0.04	82.53%
1000	**36**	**0.14**	**0.04**	**84.81%**
1000	45	0.39	0.08	79.74%
1500	35	0.18	0.15	78.48%

	H	A	S
H	41	1	0
A	4	10	3
S	2	2	16

a)

	H	A	S
H	41	1	0
A	6	8	3
S	2	1	17

b)

Fig. 7. Confusion matrices for experimented parameters with NN classifier and CF: a) parameters for the third row in Table 2; b) parameters for the first row in Table 2

For SVM classifier the polynomial and RBF kernel were tested. In Fig. 8 it is shown the results of the experiments with SVM classifier, Fig. 8a and Fig. 8b shows the confusion matrices for PF; in Fig. 8c and Fig. 8d CF are used; in Fig. 8a and Fig. 8c a polynomial kernel is applied, finally Fig. 8b and Fig. 8d show the results using a RBF kernel. The best accuracy obtained by SVM was 89.87%, which is reached when using PF as training inputs and polynomial kernel.

	H	A	S
H	42	0	0
A	1	14	2
S	2	3	15

a)

	H	A	S
H	42	0	0
A	15	0	2
S	16	0	4

b)

	H	A	S
H	42	0	0
A	11	5	1
S	0	3	18

c)

	H	A	S
H	42	0	0
A	17	0	0
S	20	0	0

d)

Fig. 8. Results with SVM classifier using as training inputs: a) PF with polynomial kernel; b) PF with RBF kernel; c) CF with polynomial kernel; d) CF with RBF kernel

Finally, Naïve Bayes classifier is used for the experiments, in Fig. 9 it can be seen the confusion matrices for the two sets of features, in Fig. 9a and Fig. 9b, PF and CF were used as input respectively. The highest accuracy (84.81%) for Naïve Bayes classifier is obtained using PF as input. Additionally, in Fig 10 the histogram of the accuracies obtained in our experiments is depicted, where it is clear that the best accuracy was obtained by NN using PF as features.

	H	A	S
H	38	4	0
A	0	11	6
S	0	2	18

a)

	H	A	S
H	36	6	0
A	5	10	2
S	0	3	17

b)

Fig. 9. Naïve Bayes classifier results: a) point features; b) coefficient features

4 Conclusions

In order to classify emotions from face images a method based on thresholding and border extraction for pre-processing stage is proposed. Other methods based on clustering algorithms such as OPTICS, CLARANS and DBscan are tested, but these approaches have not a good performance in order to get ROIs information.
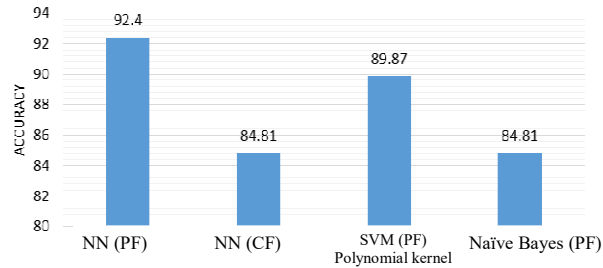
Fig. 10. Histogram with the best accuracy of experimented learners

Features such as coefficients of divided differences of Newton method and points of each ROIs was used to train the classifiers such as SVM, Naïve Bayes and Neural Networks, the best accuracy was obtained by Neural Networks using as training input the features extracted by our proposed approach. To get the accuracy of the NN some parameters were fixed such as momentum, learning rate, neurons in the hidden layer or number of epochs to get acceptable accuracy.

As future work, additional features from ROIs will be extracted in order to get a higher accuracy and less confusion in the classification stage, the six universal emotions will be classified using other learners. Finally, a database (considering the six emotions and the neutral one) will be created in order to have a database containing high quality face images. Other pre-processing method will be also explored in order to get a better performance via clustering algorithms approach.

References

1. Cohn J., Ambadar Z., Ekman P.: Observer-based measurement of facial expression with the Facial Action Coding System. In: J. A. Coan & J. J. B. Allen (Eds.), The handbook of emotion elicitation and assessment. New York, pp. 203–221 (2007)
2. Ekman P.: Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. Psychological Bulletin 115(2), pp. 268–287 (1994)
3. Corneau C., Oliu M., Cohn J., Escalera, S.: Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applicants. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(8), 1548–1568 (2016)
4. Shan C., Gong S. McOwan P.: Facial expression recognition based on Local Binary Pattern: A comprehensive study. Image and Vision Computing 27 (6), 803–816 (2009)
5. Sabu G.: Geometrical Model of Facial Regions and CUDA based Parallel Face Segmentation for Emotion Recognition. International Journal of Applied Engineering Research 11(9), 6740–6752 (2016)

6. Kim B., Roh J., Dong S., Lee S.: Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. Journal on Multimodal User Interfaces 10 (2), 173–189 (2016)

7. Sebe N., Michael S., Sun Y., Ira C., Gevers T., Thomas S.: Authentic facial expression analysi., Image and Vision Computing, 25(12), 1856–1863 (2007)

8. Pantic M., Valstar M., Rademarker R., Maat, L.: Web-based database for facial expression Analysis. In: Proceedings of International Conference on Multimedia and Expo, pp. 5–10 (2005)

9. Wen G., Hou Z., Li H., Li D., Jiang L., Xun E.: Ensemble of Deep Neural Networks with Probability-Based Fusion for Facial Expression Recognition. Cognitive Computation. 1–14 (2017)

10. Sönmez E., Cangelosi A.: Convolutional neural networks with balanced batches for facial expressions recognition. In: Proceedings of Ninth International Conference on Machine Vision, pp. 1–6 (2017)

11. Liu P., Han S., Meng Z., Tong Y.: Facial expression recognition via a boosted deep belief network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1805–1812 (2014)

12. Otsu N.: A threshold selection method from gray-level histogram. IEEE Transaction on systems, man, and cibernetics 9(1), 62–66 (1979)

13. Ester M., Kriegel H., Sander J., & Xu X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226–231 (1996)

14. Ankerst M., Breunig M., Kriegel H., Sander J.: OPTICS: ordering points to identify the clustering structure. ACM Sigmod record 28(2), 49–60 (1999)

15. Ng R., Han J.: CLARANS: A method for clustering objects for spatial data mining. IEEE transactions on knowledge and data engineering, 14(5), 1003–1016 (2002)

16. Viola P., Jones P.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Proceedings of Computer Vision and Pattern Recognition, pp. 511–518 (2001)

17. Atkinson K.: An introduction to numerical analysis, John Wiley & Sons (2008)