Supervised Methods for Classifying Facial Emotions

Francisco Aulísio dos Santos Paiva, Paula D. Paro Costa, José Mario De Martino
Department of Computer Engineering and Industrial Automation (DCA)
School of Electrical and Computer Engineering
University of Campinas (Unicamp)
Campinas, Brazil

***Abstract*—This paper presents a comparison between the K-NN (K-Nearest Neighbors) and SVM (Support Vector Machine) methods for classifying emotions. The database contains a set of 568 images of faces expressing 22 emotions. Classification is carried out in such a way as to classifying these 22 emotions as well as two other sets of categories, namely valence (positive and negative emotions) and the so-called six basic emotions (joy, sadness, fear, surprise, disgust, anger). Different sets of features were tested (statistics of histograms of regions of interest - mouth and eyes - and distances between characteristic points on the face) as well as different configurations of input parameters for training the classifiers in order to achieve the best performance. The results of the three experiments reveal accuracy values ranging from 79% to 90% for the K-NN classifier and from 88% to 94% for the SVM classifier.**

***Keywords*-Facial expression; supervised methods; valence; emotion classification;**

I. Introduction

Work by Field et al. [1] has shown that neonates can imitate and discriminate expressions of joy, sadness and surprise. Recently, Peltola et al. [2] has shown that 7-month infants are able to make cognitive associations of fear expression from a specified signal of threat. This suggests that expressing and recognizing facial emotions are part of our most primitive skills.

In fact, some studies on non-human emotions such as Plutchik [3] and the studies on facial expressions by Ekman et al. [4] collected several pieces of evidence on the existence of six basic or discrete emotions. The survey of Cornelius [5] lists happiness, sadness, fear, anger, surprise and contempt/disgust, whereas Shaver et al. [6] lists love, joy, surprise, anger, sadness and fear, based on an experimental study with three different cultures. In this paper, we only considered the list of the six basic emotions [4].

One of the main theoretical frameworks for studying emotions is the cognitivist approach (see [5] for a survey), which states that the emotional experience depends on a mechanism of appraisal by the perceiving subject. This notion of appraisal is close to the idea that emotions are tendencies to action[7]. The cognitive approach is related to several other studies on an variety of appraisal dimensions like pleasantness (valence), control, certainty, responsibility and effort, which underlies the expression of emotions. Nevertheless, Schlosberg [8] and Osgood et al. [9] have shown that only three dimensions explain most part of appraisal variance revealed by the subjects: activation, valence and attention/rejection. Valence is related to the pleasantness of a emotion, which is generally referred to as positive (pleasant) or negative (unpleasant). The proposal by Ortony, Clore and Collins [10] associated cognitive meaning to logical consequences related to the valence of an appraisal process. The so-called OCC model comprehends a set of 22 emotions: happy for, joy, hope, satisfaction, relief, pride, gratification, gratitude, admiration, love, pity, sadness, fear, resentment, fears confirmed, shame, reproach, remorse, gloating, disappointment, disgust and anger.

In this paper we present the results of three classification experiments: one using the 22 original emotions, another one using 6 basic emotions reclassified from the 22 proposed emotions and the final experiment reclassifying the original set into positive and negative emotions (valence-based only).

In a recent work by Olivera and Jaques [11] the six basic emotions were classified using a neural network, which allowed a hit rate from 63.33% to 89.87%. The database was formed by low-resolution video images. Regarding valence-based-only classification, Holkamp [12] classified valence using facial expressions of TV-viewers. It is reported a classification rate of 66% using Support Vector Machines (SVM) for a specific dataset. The work by Sohail and Bhattacharya [13] used 15 SVMs built with a kernel radial basis function to reach a classification rate from 85% to 89% depending on image resolution.

This paper aims at classifying a set of images extracted from videos of facial expressions displaying 22 different emotions. From this set, a first classification according to valence was done (positive and negative). Then, a classification in 22 emotions was carried out according to the proposal in Ortony, Clore and Collins [10]. Finally, a selection of images representing the 6 basic emotions is used for testing.

The first step was the extraction of 64 features from the facial expressions. These features were organized into three sets. First, a set of 22 features was obtained from the descriptors of pixel histograms in two regions of interest (mouth and eyes). The second set was obtained from LBP (Local Binary Pattern)-based texture features, which also formed by 22 features. Finally, 20 distances from characteristic points on the face were used as features, from the work by [14]. The sets were never used simultaneously for training. Thus, only 42 features were used for training in two experimental settings. First, 22 features of regions mentioned above plus the

20 distances, in a second experimental setting, 22 features of LBP plus 20 the distances.

The second step was the classification itself. The K-NN (k-Nearest Neighbor Learning) and SVM (Support Vector Machines) were used. The PCA (Principal Component Analysis) was used to reduce dimensionality. For feature selection, Decision Tree and Random Forest techniques were used. We compared the results accuracy values for both SVM and K-NN classifiers from the same set of features.

II. Methodology

According to Duda et al. [15] a pattern recognizer or classifier are usually built with the following stages: database collection, pre-processing, feature extraction, classification and assessment. The database is a set of samples to be classified, whereas the pre-processing is the stage where the samples can the prepared for a better performance during training. Some of the techniques at this stage are normalization, noise reduction, identification of regions of interest. The following stage, feature extraction, is a new representation of the previous data. For a better performance in classification, it is important to discard redundant or irrelevant features. This trimming operation increases generalization ability, reduces the computational complexity as well as the time during the training phase. In the following, a classifier is used for training where data is used for associating samples to categories in the most efficient way. In the last stage, assessment the performance of the classifier is measured in a test set, not previously used for training.

A. *Database of Facial Images*

A database of facial images, called CH-Unicamp, was built by Costa in [14]. CH-Unicamp is an annotated database of expressive visemes (visual phoneme images) played by a female actress whose 2D frontal face images were retained for analysis. The image frames were extracted from video recordings of the actress playing different dialogue situations. In these dialogue situations, the character played by the actress is talking with emotion to one or more imaginary characters. Each dialogue situation was designed to elicit one of the 22 emotions described by the OCC model [10].

The actress was filmed in a TV studio (HD 1920x1080 pixels, NTSC 29.97 fps), in front of chroma-key background, without markers, makeup or accessories in her head and face.

CH-Unicamp database is composed of 782 facial images (22 OCC emotions + 1 neutral expression) and the x and y-coordinates of 56 facial feature points associated to each image of the database [14]. The coordinates of the feature points were semi-automatically obtained and they were chosen to delineate the facial “shape” including the eyebrows, the eyes, the nose, the lips, the ears and the chin (Figure 1).

The database images were visually inspected and a subset of 568 images was selected for the classification experiments. Frames with neutral expressions and those that were considered transitions of expressive states (thus not clearly representing an emotion) were discarded.

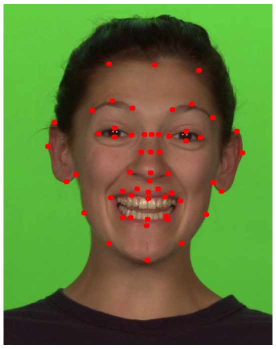

Fig. 1. Example of image from CH-Unicamp database, including its 56 feature points

B. *Pre-processing*

The pre-processing of the images is described in [14]. The facial image *shape* was defined as being the set of feature points around the contour of the eyes, the mouth, the eyebrows and the nose. By doing so, it was possible to build a vector as a result of the concatenation of all coordinates of the characteristic points of the face. From this shape vector, all images were aligned in such a way as to minimize the differences in head movement across the facial images. This alignment is necessary to obtain a better representation of the points’ distribution. The distance between characteristics points in the images were used as features in the stage of classification.

For this work, it was also necessary to convert the original color images into grayscale as well as to extract the regions of interest (ROI) shown in Figure 2. We chose to work with the regions of the mouth and the eyes because they are considered more significant to detect emotions [14].

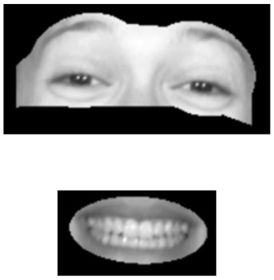

Fig. 2. Regions of Interest (ROI)

C. *Features extraction*

Three types of feature were used for feature extraction. Statistics from the pixel histograms of ROI, statistics of texture histograms in the ROI described by LBP (Local Binary Pattern) and, finally, the Euclidean distance from characteristic points on the facial image. Eleven statistical features were used, which are: mean (mean grayscale value), variance (variance of grayscale values), skewness, kurtosis, entropy, mode (grayscale value with the largest occurrence), percentile 1%, percentile 10%, percentile 50% (median of the grayscale value), percentile 90%, percentile 99%.

The Euclidean distances considered 20 distances from characteristic points on the face, such as: the distances between three points equally distributed on both eyebrows from a fixed point on the nose (6 distances), the distances defining the vertical and horizontal opening of the mouth (7 distances), the distances between the closest points in the two eyebrows to identifying frowning degree (1 distance), the distances defining the motion of the eyes (6 distances).

It is important to observe that the LBP method analyses the points around a central point by testing if pixel values are greater or lower than this central point. If greater they are changed into 1, if lower, they are changed into 0. From this transformed image, histogram of the two ROI were obtained for computing the statistical descriptors. For classifying, different combinations of these three sets of features were evaluated.

D. Classification stage

During the classification stage, the set of data is split into two subsets: a training and a test subset, for the classification tests, two types of classifiers were used: K-NN and SVM. K-NN (K-Nearest Neighbours) is a supervised classifier, which stores the training samples for reference. Classification is done by voting to the closest neighbors of a reference point. SVM (Support Vector Machine) is a classifier whose goal is to find a decision boundary between two classes, that is, to classify data from two classes by building a separating hyperplane between them. The test was performed with these two techniques because they are widely used for computer vision. Besides that, our objective is to evaluate their performance when applied to the classification of emotions in the aforementioned database.

III. Experiments

The experiments presented here were based on a comparison of the performance of the K-NN and SVM classifiers. Both were tested with different combinations of the extracted features, as shown in the results section. For both classifiers we used the Grid Search function. This function searches for the best parameters for the classifiers. This means that the algorithm of the classifiers runs for each set of parameters for allowing the grid search function to choose the best estimator. For instance, the K-NN classifier was tested with the following set of parameters: number of neighbors: $[1, 2, 3, 4, 5, 6, 7]$; the metric for computing the distance was the Minkowski one (L1-norm and L2-norm). As for the SVM classifier, the penalty parameter was tested with values of 1 and 10, whereas the kernel tested were linear, poly and rbf. It is worth mention that the grid search was used for the set Stratified ShuffleSplit, which is a crossvalidation technique for which the data set is split randomly into test and training subsets using the respective proportion of 30% and 70% for the first two experiments and the respective proportion of 20% and 80% for the third experiment. This option makes less probable an inappropriate split. Explanations of these parameters can be found in [16].

The first experiment tested both classifiers (K-NN and SVM) for splitting the categories according to valence into positively and negatively-valenced images. From the best result in this first experiments, two other experiments tested the classification of 22 emotions according to the OCC model as well as the six basic emotions proposed by Ekman.

IV. Results and Discussion

The first experiment was made from the division of database in valences (positive, negative). In the following table, the accuracy values of all tests carried out for the valence classification are shown. The best parameters (neighbor and norm) for the K-NN classifier are: $k = 2$, $L1$; $k = 2$, $L2$; $k = 2$, $L2$; $k = 4$, $L2$; $k = 4$, $L2$; $k = 6$, $L2$; $k2$, $L2$; $k = 4$, $L2$; $k = 2$, $L2$; $k2$, $L1$. These pairs of values follow the same order found in TABLE I. It is possible to infer that the SVM classifier was the best classifier in all tests carried out. Furthermore, the use of distances between characteristic points in the face had a significant role in the classification performance.

TABLE I
CLASSIFICATION RESULTS

Accuracy		
Features	**K-NN**	**SVM**
Eyes	82%	88%
Lips	79%	84%
Eyes + Lips	82%	89%
Distances	88%	93%
All	89%	94%
LBP	84%	86%
LBP + Dist.	87%	93%
PCA	90%	93%
Decision Tree	89%	93%
Random Frst.	87%	93%

In fact, by using only distances as features in the SVM classifier, an accuracy of 93% was obtained, whereas, by using all the features combined (histogram descriptors of eyes and lips + distances) only a 1% gain is obtained. The use of the techniques of PCA, Decision tree and Random Forest did not have an effect on the accuracy by using the SVM classifier. The confusion matrix of this experiment is $\begin{pmatrix} 83 & 4 \\ 8 & 67 \end{pmatrix}$. The matrix shows that the method hit 83 images as positive, but misclassified 4 images, classifying them as negative. Also it hit 67 negative images, but it misclassified 8 as positive.

In Figure 3, the left image (admiration) was classified as negative and the right image (anger) was classified as positive. A possible reason for misclassification is the fact that the former image has the eyes and the mouth closed and the second is showing the teeth, which resembles a smile. Since a 93% accuracy was obtained by using the distances only, we retained this set of features only to perform the classification

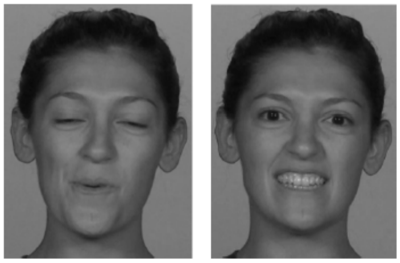

Fig. 3. Cases of misclassification.

of the 22 emotions of the OCC model and the classification of the 6 basic emotions.

Thus, the second experiment for classifying 22 emotions allowed us to obtain a 34% value for the accuracy. In this case, we observed high levels of confusion among emotion expressions such as as happy-for, joy, satisfaction and hope or anger and disgust.

For the third experiment, with six basic emotions, we used the expression of admiration instead of surprise, for the lack of the latter in the database and the closeness of the admiration and surprise expressions. The obtained accuracy for this case was 84%, using the distance as only features, as said previously. By analyzing the errors, we verified 5 misclassification, where admiration was classified as sadness and disgust, joy, as admiration, sadness as disgust and disgust as anger. When admiration is removed from the categories to be classified, we got 94% of accuracy for this set of five emotions, where only 2 misclassification appear. One expression of sadness was misclassified as joy, and one expression of disgust was misclassified as anger.

V. Conclusion

This work presented an exploratory study regarding the problem of classifying the valence and the emotion of expressive speech facial images, through the use of supervised classification methods.

Two methods for classifying emotions were analyzed in this work using facial expression images extracted from video recordings of a female actress [14]. The K-NN and SVM methods were used to classify emotional categories according to valence, the six classic, basic emotions and the 22 emotions of the OCC model. For doing so, only the regions around the eyes and mouth were used to extract a set of 22 texture-related features and their LBP-modified counterparts. Additionally distances between characteristic points on the face for characterizing modifications in the eyebrows, eyes and mouth expressions were used.

It was possible to conclude that the use of distances between characteristic points on the face provided promising results for the classification of valence and emotions, given a small set of stereotypical emotions. However, the discrimination of complex emotions, with subtle differences in facial expressions is a challenging problem.

This result is compatible with the literature reviewed above, although it is not easy to compare with, given differences in databases. The verification of which expressions were misclassified suggests the kind of generalization the automatic method is performing. In one of the anger expressions, the teeth are shown and the eyebrows are raised, which was predicted as a positive emotion.

Future work includes the study of other faces' regions for improving such classification as well as the analysis of sequential frames to recognize the emotion label of an expressive speech video excerpt.

References

[1] T. M. FIELD, R. WOODSON, and e. a. COHEN, "Discrimination and imitation of facial expressions by term and preterm neonates," *American Association for the Advancement of Science*, vol. 218, no. 4568, pp. 179–181, 1982.

[2] M. J. PELTOLA, J. M. LEPPÄNEN, S. MÄKI, and J. K. HIETANEN, "Emergence of enhanced attention to fearful faces between 5 and 7 months of age," *Social Cognitive and Affective Neuroscience*, p. nsn046, 2009.

[3] R. PLUTCHIK, "Emotion: a psychoevolutionary synthesis," *New York: Harper and Row*, 1980.

[4] P. EKMAN, W. V. FRIESEN, and et al., "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.

[5] R. R. CORNELIUS, *The science of emotion: Research and tradition in the psychology of emotions.* Prentice-Hall, Inc, 1996.

[6] P. SHAVER, J. SCHWARTZ, D. KIRSON, and C. O'CONNOR, "Emotion knowledge: further exploration of a prototype approach." *Journal of personality and social psychology*, vol. 52, no. 6, p. 1061, 1987.

[7] N. H. FRIJDA, *The Emotions.* Cambridge: Cambridge university Press, 1986.

[8] H. SCHLOSBERG, "Three dimensions of emotion." *Psychological review*, vol. 61, no. 2, p. 81, 1954.

[9] C. E. OSGOOD, *The measurement of meaning.* Urbana: Univ. of Illinois, 1957.

[10] A. ORTONY, G. L. CLORE, and A. COLLINS, *The cognitive structure of emotions*. Cambridge university press, 1988.

[11] E. OLIVEIRA and P. A. JAQUES, "Classificação de emoções básicas através de imagens capturadas por webcam," *Revista Brasileira de Computação Aplicada*, vol. 5, no. 2, pp. 40–54, 2013.

[12] Y. H. HOLKAMP, "Classification of valence using facial expressions of tv-viewers," Ph.D. dissertation, TU Delft, Delft University of Technology, 2014.

[13] A. SOHAIL and P. BHATTACHRYA, "Classifying facial expressions using point-based analytic face model and support vector machines," in *2007 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2007, pp. 1008–1013.

[14] P. COSTA, "Two-dimensional expressive speech animation," Ph.D. dissertation, UNICAMP, 2 2015, Campinas, SP, 2015.

[15] R. DUDA, P. HART, and D. G. STORK, *Pattern classification*. John Wiley & Sons, 2012.

[16] C. CHANG and C. LIN, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.