FACIAL EXPRESSION RECOGNITION USING ENSEMBLE OF CLASSIFIERS

T. H. H. Zavaschi, A. L. Koerich[*]

Pontifical Catholic University of Paraná
Department of Computer Science
Curitiba, PR, Brazil

L. E. S. Oliveira[†]

Federal University of Paraná
Department of Computer Science
Curitiba, PR, Brazil

ABSTRACT

This paper presents a novel method for facial expression classification that employs the combination of two different feature sets in an ensemble approach. A pool of base classifiers is created using two feature sets: Gabor filters and local binary patterns (LBP). Then a multi-objective genetic algorithm is used to search for the best ensemble using as objective functions the accuracy and the size of the ensemble. The experimental results on two databases have shown the efficiency of the proposed strategy by finding powerful ensembles, which improves the recognition rates between 5% and 10%.

Index Terms— Face recognition, Emotion recognition.

1. INTRODUCTION

Automatic facial expression recognition has been subject of investigation in the last years due to the great number of potential day-to-day application such as human-computer interaction, emotion analysis, operator fatigue detection in industries, interactive video, indexing and retrieval in video databases, image understanding, and synthetic face animation. Facial expression recognition is also a necessary step towards a computer facilitated human interaction system as facial expressions play a significant role in conveying human emotions [1]. Due to such an importance, a lot of effort has been devoted to build reliable automatic facial expression recognition systems. The methods reported in the literature can be classified basically into geometry analysis and appearance-based. The former takes into account some predefined geometric positions, also known as fiducial points, as facial features to represent facial expressions [2]. However, the geometric feature-based representation commonly requires accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations [3]. The second approach models the appearance changes of the faces through a holistic spatial analysis. Among the tools used for this approach are Principal Component Analysis [4], Independent Component Analysis [5], Gabor filters [6], and LBP [7]. According to the literature, Gabor filters yield superior performance for facial analysis and for this reason they have been widely adopted [6, 8, 9]. The downside, though, is the elevated computational cost in terms of time and memory usage. Recently LBP have been introduced as effective appearance features for facial image analysis [3, 10]. Experiments have demonstrated that compared with Gabor filters, LBP features save much computational resource whilst retaining facial information efficiently [3].

Though much progress has been made, recognizing facial expressions with a high accuracy remains difficult due to the subtlety, complexity, and variability of facial expressions. An efficient way to deal with complex pattern recognition problems such as face expression recognition is to build ensemble of classifiers to take advantage of the inherent diversity introduced by classifiers trained with different feature sets. Several studies have been published demonstrating the benefits of the combination paradigm over the individual classifier models. During the last years, a considerable amount of research has gone into ensemble of classifiers. The effectiveness of such methods comes primarily from the diversity caused by resampling the training set while using the complete set of features to train the component classifiers. In addition, some attempts have been made to incorporate the diversity into ensemble creation methods by over-producing classifiers and then choosing some of them to compose the ensemble. An alternative to bring diversity to the ensemble is to combine classifiers trained with different feature sets. The efficiency of this strategy has been reported by several authors [11].

In this work we propose an ensemble of classifiers based on the under-pinning concept of "over-produce and choose". The pool of base classifiers is created using the two more prominent feature sets used for facial expression recognition, namely, Gabor Filters and LBP. Then a multi-objective genetic algorithm is used to search for the best ensemble using as objective functions the accuracy and size of the ensemble. Through a set of comprehensive experiments on two different databases we demonstrate the efficiency of the proposed strategy by finding powerful ensembles, which succeed in improving the recognition rates from 5% to 10%. The results reported in this paper compare favorably to the literature.

This paper is organized as follows: Section 2 outlines the proposed methodology to create ensemble of classifiers. Section 3 introduces the feature sets used to train the pool of base classifiers. The experimental results are presented in Section 3. Finally, conclusions are stated in the last section.

2. METHODOLOGY

The approach proposed to generate ensemble of classifiers is based on an "overproduce and choose" paradigm where a pool of classifiers is created by varying parameters of Gabor filters and LBP operators. Once this pool of classifiers have been trained, the second level is suggested to choose the members of the team which are small (few classifiers) and accurate. The second level can be performed by any search algorithm. We have chosen a multi-objective genetic algorithm (MOGA) to such an aim because building an ensemble of classifiers can be formulated as a multi-objective problem since we want to minimize not only the error rate of the ensemble but also

[*]Also with Department of Electrical Engineering at Federal University of Paraná. This research has been supported by CNPq grant 309.295/2007-6

[†]This research has been supported by CNPq grant 471.496/2007-3

the number of the classifiers in the ensemble. In this context, MOGAs are more suitable than single genetic algorithms (GA) because they can provide a set of solutions known as Pareto-optimal. Single GA, on the other hand, converge to a specific region of the search space depending on the weights assigned for each objective. More details about the limitations of the single GA for multi-objective optimization problems can be found in [12]. In this context, let $A = C_1, \ldots, C_L$ be a set of L classifiers and B a chromosome of size L of the population. The gene i of the chromosome B is represented by the classifier C_i from A. Thus, if a chromosome has all bits selected, all classifiers of A will be included in the ensemble. Fig.1 depicts the proposed methodology.

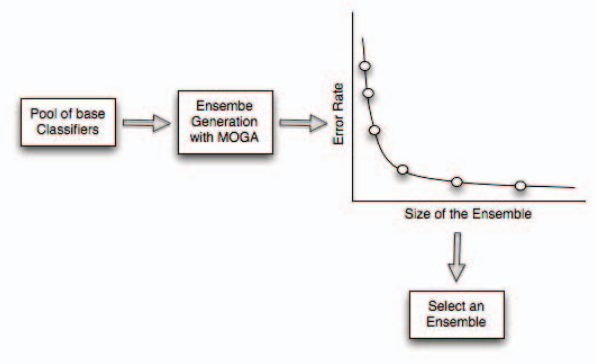

Fig. 1. The overview of the methodology

We selected as objective to be optimized the accuracy and the size of the ensemble regardless the diversity of the classifiers because of the nature of the application. Since facial expression recognition usually is applied to on-line systems, performance is a crucial requirement that this kind of application should meet. Therefore smaller ensembles appear more suitable in this case.

3. CLASSIFIER AND FEATURE SETS

The overproduce stage is done by varying parameters of both Gabor filters and LBP operators. Both feature set have been successfully applied to facial expression recognition [3, 13] and for this reason they were selected to train our base classifiers. All the classifiers used in this work are Support Vector Machines (SVM) trained with Gaussian kernel. Kernel parameters such as C and γ were defined through a grid search using k-fold cross validation.

3.1. Gabor Filters

A family of Gabor kernel is the product of a Gaussian envelope and a plane wave, as defined in Eq.1

$$\Psi_{u,v}(z) = \frac{||k_{u,v}||^2}{\sigma^2} e^{-||k_{u,v}||^2/\sigma^2} [e^{ik_{u,v}z} - e^{-\sigma^2/2}] \quad (1)$$

In this case, $z = (x, y)$ is the variable in the spatial domain and $k_{u,v}$ (Eq.2) is the frequency vector, which determines the scales and orientations of Gabor kernels.

$$k_{u,v} = \frac{k_{max}}{f^v} e^{i\Phi_u} \quad (2)$$

where $k_{max} = \frac{\pi}{2}$, $f = \sqrt{2}$, and $\Phi_u = \frac{u\pi}{8}$, where u and v are orientation and scale factors, respectively. By varying u and v we can select different kernels.

Given an image $I(z)$, its Gabor transformation at a particular position can be computed by a convolution with Gabor Kernels using Eq.3.

$$G_{u,v} = I(z) \times \Psi_{u,v}(z) \quad (3)$$

The magnitude of the resulting complex image is given by Eq.4.

$$|G| = \sqrt{\Re\mathfrak{e}(G)^2 + \Im\mathfrak{m}(G)^2} \quad (4)$$

All features derive from $|G|$ and the feature vector $F_{k,N}$ is given by Eq.5

$$F_{k,l} = \sum_{i=x_{l-k}}^{x_{l+k}} \sum_{j=y_{l-k}}^{y_{l+k}} |G_{i,j}|, l = 0, 1, \ldots, N; k = 0, 1, \ldots 5. \quad (5)$$

where N is the number of the fiducial points marked in the face image. Koutlas and Fotiadis [13] proposed a set of 20 fiducial points which were derived from 74 different landmarks (Fig.2). According to the authors, such points lie around prominent features of the face that contain the most significant information regarding the muscle movement which is responsible for facial expressions.

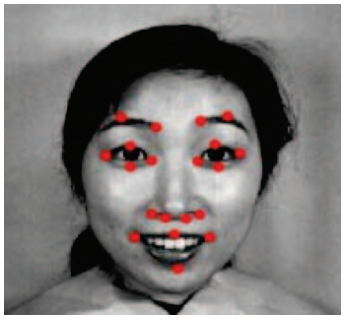

Fig. 2. The 20 fiducial points proposed by Koutlas and Fotiadis [13].

For each fiducial point a mask of size $k \times k$ is used to compute the feature vector according to Eq.5, where $k = \{1, 3, 5, 7, 9\}$. We extracted five feature sets based on scales with 160 components each, eight feature sets based on orientations with 100 components each, and one feature set with 800 components combining scales and orientations. Considering the five different masks, we have 70 different feature sets that will be used to train 70 different base classifiers.

3.2. Local Binary Patterns (LBP)

The operator $\text{LBP}_{P,R}$ produces 2^P different binary patterns that can be formed by the P pixels in the neighbor set. (P, R) stands for a neighborhood of P equally spaced sampling points on a circle of radius of R that from a circularly symmetric neighbor set. However, certain bins contain more information than others, hence, it is possible to use only a subset of the 2^P LBPs. Those fundamental patterns are known as uniform patterns. A LBP is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 001110000 and 11100001 are uniform patterns. It is observed that uniform patterns account for nearly 90% of all patterns

in the (8,1) neighborhood and for about 70% in the (16,2) neighborhood in texture images [7].

Accumulating the patterns which have more than two transitions into a single bin yields an LBP operator, denoted $\text{LBP}^{u2}_{P,R}$, with less than 2^P bins. For example, the number of labels for a neighborhood of 8 pixels is 256 for the standard LBP but 59 for LBP^{u2}. Thereafter, a histogram of the frequency of the different labels produced by the LBP operator can be built. According to Shan et al. [3], an interesting way of using LBP in face images consists in equally divide the image into n small zones $Z_0, \ldots, Z_n$ to extract the LBP histograms. The features extracted from each zone are then concatenate into a single vector.

In our approach the faces were divided into 42 zones (7×6). Three different configurations of the LBP operator were considered: $\text{LBP}^{u2}_{8,1}$, $\text{LBP}^{u2}_{8,2}$, $\text{LBP}^{u2}_{16,2}$. The first two produce a feature vector of 59 components per zone, summing up 2,478 components while the last one produces a feature vector of 243 components per zone, summing up, 10,206 components.

4. EXPERIMENTAL RESULTS

Two databases were used in the experiments: JAFFE database [1] and Cohn-Kanade database. The JAFFE database contains 10 female individuals and 213 images of facial expressions. Each image has a resolution of 256×256 pixels. The number of images corresponding to each of the 7 categories of expression (neutral, happiness, sadness, surprise, anger, disgust and fear) is almost the same. Each image in the database was rated by 91 experimental subjects for degree of each of the six basic expressions present in the image [17]. The Cohn-Kanade database consists of image sequences depicting the evolvement of every facial expression from the neutral state until it reaches its highest intensity in the last frame. The database is encoded into combinations of Action Units. These combinations were translated into facial expressions to define the corresponding ground truth for the facial expressions [14]. All subjects were taken under consideration to form the database, composed of 1,281 images.

Two different experiments were performed in each database. In Experiment (I) individuals that participate in the training set could be part of the testing set. Of course that those images used for training were not used for testing. In Experiment (II), the individuals used for training were not used for testing. Due to the small size of the public datasets used for this kind of research, the first approach is very often found in the literature. However, the second case is far more realistic since during the deployment phase the system would have to classify expressions from people not used to train the system.

The first step of the proposed methodology consists in training the pool of 73 base SVM classifiers. The classifiers are separated into three groups: 3 LBP, 30 Gabor scale-based, and 40 Gabor orientation-based classifiers. In all experiments we have used 10-fold cross validation similar with that in [9]. After training the pool of classifiers they are used as input to the MOGA. In this work we have used the Non-Dominated Sorting Genetic Algorithm II (NSGA II) proposed by Deb et al. [12]. The idea behind the NSGA is that a ranking selection method is used to emphasize good points and a niche method is used to maintain stable subpopulations of good points. Before the selection is performed, the population is ranked based on an individual's non-domination. The non-dominated individuals present in the population are first identified from the current population. Then, all these individuals are assumed to constitute the first non-dominated front in the population and assigned a large dummy fitness value. The same fitness value is assigned to give an equal reproductive potential to all these non-dominated individuals.

In our experiments, the NSGA is based on bit representation, one-point crossover, bit-flip mutation, and roulette wheel selection (with elitism). The following parameters were employed: population = 100, number of generations = 300, probability of crossover = 0.7, probability of mutation = 0.01, and niche distance = 0.05. The size of the chromosome is 73, since we have 73 classifiers. The error rate of the ensemble is computed through the Sum rule since this was the rule that produced better results. To define the probabilities of crossover and mutation, we have used the one-max problem, which is probably the most frequently used test function in research on genetic algorithms because of its simplicity. The population size and the number of generations were defined empirically. The evolution of the population in the objective plane for Experiments (I) and (II) reveal that in both cases the algorithm converges toward the Pareto-front producing a set of possible solutions.

The next step consists in choosing the best ensemble of classifiers from the Pareto. High accuracy is important but the size of the ensemble is also an important issue for this kind of application. The ensembles that provide the best trade-off between accuracy and size are located close to the end of the Pareto. The selected classifiers and their individual performances are reported in Tab.1. The selected ensembles were present in all the 10 replications, what guarantee that the ensembles were not found accidentally.

Table 1. Selected Classifiers - JAFFE Database

Experiment I		Experiment II	
Feature Set	Acc. (%)	Feature Set	Acc. (%)
$\text{LBP}_{8,2}$	87.3	$\text{LBP}_{8,2}$	60.6
Gabor S:5 M:3×3	91.6	$\text{LBP}_{16,2}$	59.3
Gabor O:3 M:7×7	80.7	Gabor O:2 M:1×1	41.0
Gabor O:6 M:7×7	76.6	Gabor O:3 M:5×5	41.8
Gabor O:8 M:7×7	85.9	Gabor O:6 M:9×9	41.2
All Classifiers	92.5	All Classifiers	49.0
Ensemble	96.2	Ensemble	70.0

In spite of the same size, the composition of the ensemble in the two experiments is totally different, with the exception of the LBP classifier $\text{LBP}_{8,2}$. Tab.1 shows that the problem of Experiment (II) is quite more difficult than the problem of Experiment (I). In the case of Experiment (I), the ensemble brought an improvement of about 5% compared to the best classifier. A more impressive improvement, though, was achieved in Experiment (II) where the ensemble improves the recognition rate in about 10%. A quick look on the performance of the selected classifiers for Experiment (II) would suggest that we could discard the three Gabor-based classifiers since they have a poor performance when compared with the LBP-based classifiers. In spite of the poor performance, these weak classifiers are still very important since they provide complementary information which is crucial for the good performance of the ensemble. By removing the three Gabor-based classifiers the performance of the ensemble would drop to 62%.

Tab.2 shows the performance of different approaches reported in the literature on JAFFE database. All works have used the protocol that we have employed in Experiment (I). Some of these results are not comparable directly as some authors exclude some classes of the problem. In spite of this fact, the proposed methodology compares favorably to the literature.

The same protocol used for JAFFE database was applied on the Cohn-Kanade database. However, such a database is less complex

Table 2. Comparison with different approaches on JAFFE Database

Reference	Acc. (%)	Features
[9]	90.1	Geometry and Gabor
[11]	92.5	Gabor filters
[13]	92.3	Gabor filters
[15]	94.5	LBP, Tsallis Entrop., Global App.
[16]	95.9	2D Locality Preserving Projec.
[17]	90.2	Gabor and LVQ
Proposed Appr.	96.2	Ensemble Gabor and LBP

than the JAFFE database since the facial expression images were extracted from video sequences which reduces considerably the variability of the same individual. This explains the compelling performance of some classifiers, especially in Experiment (I) where the same individual participates in both training and testing sets. Here the algorithm also converges toward the Pareto-front producing a set of possible solutions. The selected classifiers and their individual performances are reported in Tab.3.

Table 3. Selected Classifiers - Cohn-Kanade Database

Experiment I		Experiment II	
Feature Set	Acc. (%)	Feature Set	Acc. (%)
$LBP_{8,2}$	99.0	$LBP_{8,2}$	84.3
Gabor S:6 M:1×1	98.7	Gabor S:1 M:7×7	78.7
All Classifiers	98.3	All Classifiers	79.2
Ensemble	99.4	Ensemble	88.9

Since this dataset is less complex than the previous one, it requires smaller ensembles to reduce the overall error rates. In both cases, the best classifier ($LBP_{8,2}$) was selected together with a Gabor scale-based classifiers. Differently from the Experiment (I) where a single classifier almost reached the upper-limit in terms of correct classification (99%), in the Experiment (II) we got an improvement of more than 4% compared to the best classifier. This corroborates to our previous findings that weaker classifiers can bring important information to the ensemble.

5. CONCLUSION

We have described a methodology for ensemble creation underpinned on the paradigm "overproduce and choose". The pool of base classifiers is created by varying the parameters of two feature sets widely used for automatic facial expression recognition, Gabor filters and LBP. After training the pool of 73 classifiers they are used as input to an efficient search algorithm which returns a set of possible ensembles. The size and accuracy of the ensemble were the objective functions used to guide the search.

The feasibility of the strategy was demonstrated through comprehensive experiments carried out on two different databases using two different experimental protocols. The results attained demonstrated the efficiency of the proposed strategy by finding powerful ensembles, which succeed in improving the recognition rates from 5 to 10%. Such results compare favorably to the results reported in the literature.

6. REFERENCES

[1] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *IEEE Int'l Conf. on Autom. Face and Gesture Recog.*, pp. 200–205, 1998.

[2] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.27, no.5, pp.1–16, 2005.

[3] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Comput.*, vol.27, pp.803–816, 2009.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. of Cognitive Neuroscience*, vol.1, pp.71–86, 1991.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.19, no.7, pp.711–720, 1997.

[6] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.21, no.12, pp.1357–1362, 1999.

[7] T. Ojala, M. Pietikinen, and T. Menp, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, no.7, pp.971–987, 2002.

[8] S. Xiea, S. Shana, X. Chena, X. Mengc, and W. Gao, "Learned local Gabor patterns for face representation and recognition," *Signal Process.*, vol.89, no.12, pp.2333–2344, 2009.

[9] Z. Zhang, M. J. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *IEEE Int'l Conf. on Autom. Face and Gesture Recog.*, pp.454–459, 1998.

[10] C. Shan, S. Gong, and P.W. McOwan, "Robust facial expression recognition using local binary patterns," in *IEEE Int'l Conf. on Image Proc.*, pp.370–373, 2005.

[11] W. Liu and Z. Wang, "Facial expression recognition based on fusion of multiple Gabor features," in *Int'l Conf. on Patt. Recog.*, pp.536–539, 2006.

[12] K. Deb, A. Agarwal, T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Trans. Evol. Comput.*, vol.6, no.2, pp.181–197, 2002.

[13] A. Koutlas and D. I. Fotiadis, "An automatic region based methodology for facial expression recognition," in *IEEE Int'l Conf. on Syst. Man and Cybernetics*, pp.662–666, 2008

[14] M. Pantic and L. J. M. Rothkrantz, "Expert system for automatic analysis of facial expressions," *Image and Vision Comput.*, vol.18, no.11, pp.881–905, 2000.

[15] S. Liao, W. Fan, C. S. Chung, and D.-Y. Yeung, "Facial expression recognition using advanced local binary patterns," in *Int'l Conf. on Image Proc.*, pp.665–668, 2006.

[16] R. Zhi and Q. Ruan, "Facial expression recognition based on two-dimensional discriminant locality preserving projections," *Neurocomputing*, vol.71, pp.1730–1734, 2008.

[17] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Eng. Appl. of Artif. Intell.*, vol.28, pp.1056–1064, 2008.