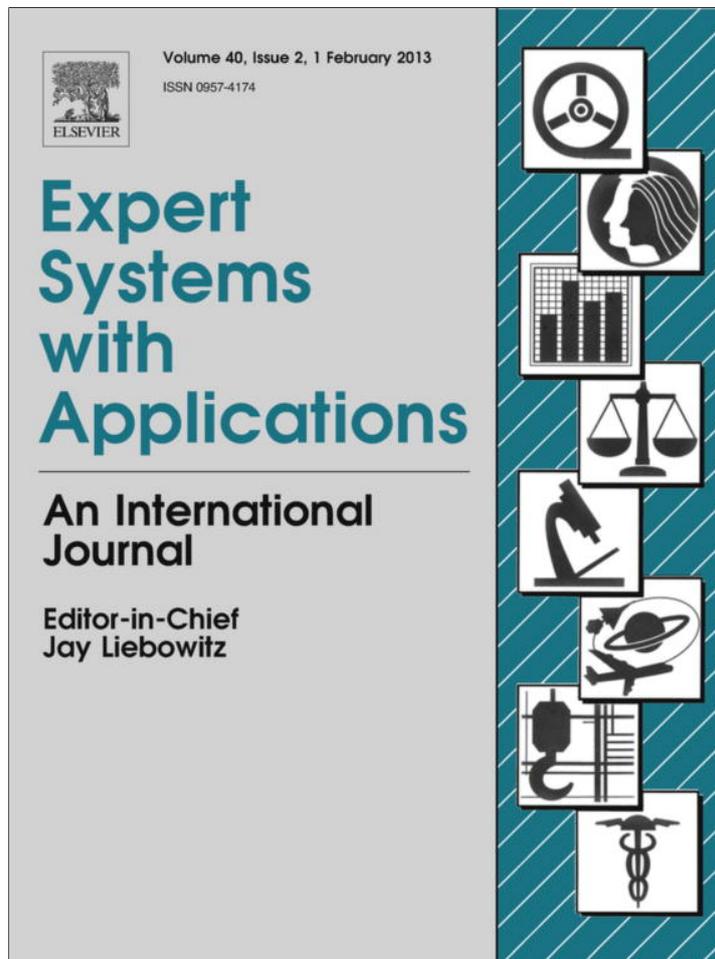


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Fusion of feature sets and classifiers for facial expression recognition

Thiago H.H. Zavaschi^a, Alceu S. Britto Jr.^a, Luiz E.S. Oliveira^b, Alessandro L. Koerich^{a,b,*}^a Pontifical Catholic University of Paraná (PUCPR), R. Imaculada Conceição, 1155, Curitiba, PR 80215-901, Brazil^b Federal University of Paraná, R. Cel. Francisco H. dos Santos, 100, Curitiba, PR 81531-990, Brazil

ARTICLE INFO

Keywords:

Face recognition
Emotion recognition
Ensemble of classifiers
Feature selection

ABSTRACT

This paper presents a novel method for facial expression recognition that employs the combination of two different feature sets in an ensemble approach. A pool of base support vector machine classifiers is created using Gabor filters and Local Binary Patterns. Then a multi-objective genetic algorithm is used to search for the best ensemble using as objective functions the minimization of both the error rate and the size of the ensemble. Experimental results on JAFFE and Cohn-Kanade databases have shown the efficiency of the proposed strategy in finding powerful ensembles, which improves the recognition rates between 5% and 10% over conventional approaches that employ single feature sets and single classifiers.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Automatic facial expression recognition has been a subject of investigation in the last years due to the great number of potential day-to-day applications such as human–computer interaction (HCI), emotion analysis, automated tutoring systems, smart environments, operator fatigue detection in industries, interactive video, indexing and retrieval of image and video databases, image understanding, and synthetic face animation (Aleksic & Katsaggelos, 2006). Furthermore, automatic facial expression recognition systems can provide a less intrusive method to apprehend the emotion activity of a person of interest (Bashyal & Venayagamoorthy, 2008). As pointed out by Lyons, Akamatsu, Kamachi, and Gyoba (1998), 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. Any natural HCI system thus should take advantage of the human facial expressions.

In 1971, Ekman and Friesen (1971) postulated six primary emotions that possess each a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as so called basic emotions. They seem to be universal across human ethnicity and cultures and comprise happiness, sadness, fear, disgust, surprise and anger. Due to the advancements accomplished in related research areas such as face detection and recognition in the beginning of the 90s, researchers renewed the interest for facial expression recognition (Fasel & Luettnin, 2003).

* Corresponding author. Address: Pontifical Catholic University of Paraná (PUCPR), R. Imaculada Conceição, 1155, Curitiba, PR 80215-901, Brazil.

E-mail addresses: zavaschi@ppgia.pucpr.br (T.H.H. Zavaschi), alceu@ppgia-pucpr.br (A.S. Britto Jr.), lesoliveira@inf.ufpr.br (L.E.S. Oliveira), alekoe@computer.org (A.L. Koerich).

A pioneering work in this field was presented by Mase and Pentland back in 1991 (Mase & Pentland, 1991).

Since then a lot of effort has been made to build more reliable automatic facial expression recognition. 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 (Besinger et al., 2010; Geetha, Ramalingam, Palanivel, & Palaniappan, 2009; Pantic & Patras, 2006; Wong & Cho, 2009; Zhang & Ji, 2005). However, the geometric feature-based representation commonly requires accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations (Shan, Gong, & McOwan, 2009).

The second approach models the appearance changes of the faces through an holistic spatial analysis. Among the tools used for this approach are Principal Component Analysis (PCA) (Turk & Pentland, 1991), Independent Component Analysis (Belhumeur, Hespanha, & Kriegman, 1997), Gabor filters (Lyons, Budynek, & Akamatsu, 1999; Tan & Triggs, 2007), and Local Binary Patterns (LBP) (Ojala, Pietikainen, & Harwood, 1996; Tan & Triggs, 2007). According to the literature, Gabor filters lead superior performance for facial analysis and for this reason they have been widely adopted (Lyons et al., 1999; Xiea, Shana, Chena, Mengc, & Gao, 2009; Zhang, Lyons, Schuster, & Akamatsu, 1998). 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 (Shan, Gong, & McOwan, 2005, 2009; Tan & Triggs, 2007). Experiments have demonstrated that when compared with Gabor filters, the simple LBP features save much computational resource whilst retaining facial information in an efficient way (Shan et al., 2009; Zavaschi, Oliveira, & Koerich, 2011).

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, which is the case of face expression recognition, is to build ensemble of classifiers to take advantage of the inherent diversity introduced by classifiers trained with different feature sets (Zavaschi et al., 2011). Several studies have been published demonstrating the benefits of the combination paradigm over the individual classifier models (Kuncheva, 2004). During the last years, a considerable amount of research has gone into ensemble of classifiers. According to the literature, the most popular methods for ensembles creation are Bagging (Breiman, 1996), Boosting (Freund & Schapire, 1996) and Random Subspaces (Ho, 1998). The effectiveness of such methods comes from the diversity caused by re-sampling the training set or even by varying the subset 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. In this direction, an interesting alternative to bring diversity to the ensemble is to combine classifiers trained with different feature sets. The efficiency of such a strategy has been reported by several authors (Kittler, Hatef, Duin, & Matas, 1998; Liu & Wang, 2006; Oza & Tumer, 2008).

In this paper 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 currently 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 the size of the ensemble. Two different experimental protocols were employed to evaluate the proposed approach. In the first one the subjects can be part of both the training and testing set (not with the same images) while in the second experiment the subjects used for training are not included in the testing set. The first protocol is very often used in the literature due to the small size of the public datasets, however, the second protocol seems to be more realistic since during the deployment phase the system would have to classify expressions from people never seen by the system.

Through a set of comprehensive experiments on two different databases (JAFFE and Cohn-Kanade) we demonstrate the efficiency of the proposed strategy by finding powerful ensembles, which succeed in improving the recognition of facial expression from 5% to 10% when compared to conventional approaches that employ

single feature vectors and single classifiers. Furthermore, the results reported in this paper compare favorably to other results found in 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 4. Finally, conclusions are stated in the last section.

2. Methodology overview

In this section we outline the approach proposed to generate ensemble of classifiers for automatic facial expression recognition which is based on a two step paradigm: “overproduce and choose” which is depicted in Fig. 1. At the first step, a pool of classifiers is created by varying the parameters of Gabor filters – orientation and scale – as well as the parameters of the LBP operators – number of points and radius of a circular mask. Once this pool of classifiers has been trained, at the second step is suggested to choose the members of the team which are small (few classifiers) and accurate (few errors). The second step can be performed by any search algorithm.

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 the number of the classifiers in the ensemble. In this context, multi-objective genetic algorithms (MOGA) 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 Deb (2001). In this work we have used the Non-Dominated Sorting Genetic Algorithm II (NSGA II) to build an ensemble of classifiers while minimizing both the error rate and the number of classifiers of the ensemble (Deb, Agarwal, & Meyarivan, 2002).

The idea behind the NSGA II is that a ranking selection method is used to emphasize good points and a niche method is used to maintain stable subpopulation of good points. It differs from simple GA only in the way the selection operator works. The crossover and mutation remain as usual. 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

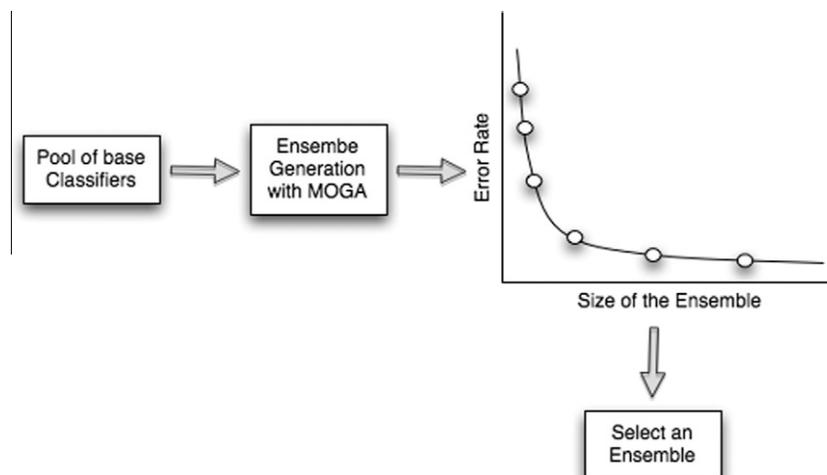


Fig. 1. The overview of the proposed method to generate ensemble of classifiers.

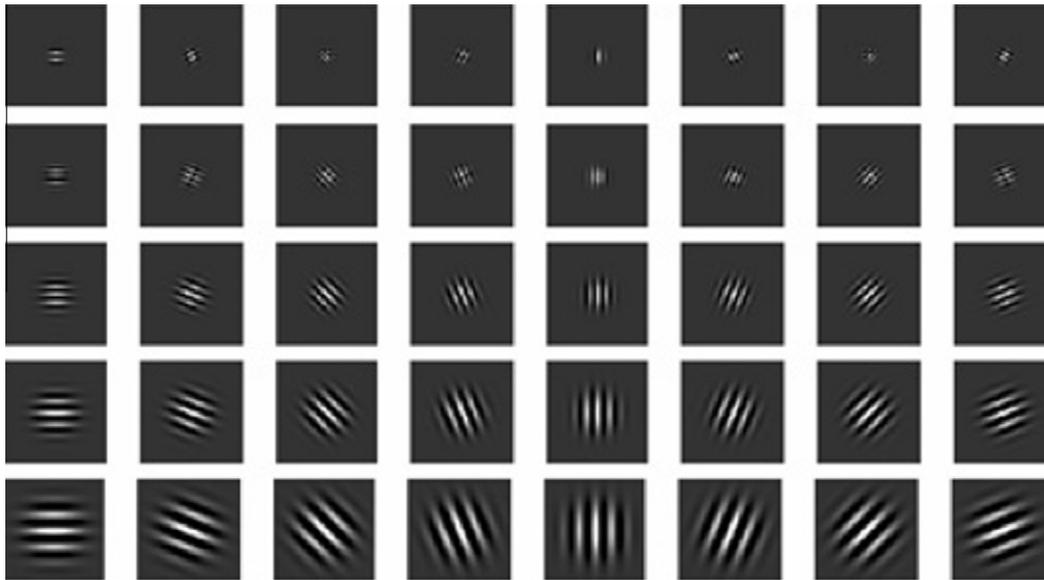


Fig. 2. Example of the Gabor filters for 8 orientations (columns) and 5 scales (rows).

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. More details about NSGA II can be found in Deb et al. (2002).

When discussing ensembles of classifiers one could argue that diversity of the classifiers is one objective that should be considered (Santos, Sabourin, & Maupin, 2009). We agree with that, but in this work we selected as objective to be optimized the accuracy and the size of the ensemble 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.

Let $A = \{C_1, C_2, \dots, C_L\}$ be a set of L classifiers and B a chromosome of size L of the population. The relationship between A and B is straightforward, i.e., 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.

3. Feature sets

This section presents the feature sets that have been chosen to model the facial expressions.

3.1. Gabor filters

Gabor filters have been successfully applied to facial expression recognition (Koutlas & Fotiadis, 2008) and for this reason they were chosen as one feature set used to train our base classifiers. A family of Gabor kernel is the product of a Gaussian envelope and a plane wave, as defined in Eq. (1)

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

where $z = (x, y)$ 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.

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

where $k_{max} = \frac{\pi}{2}$, $f = \sqrt{2}$, and $\Phi_\mu = \frac{\mu\pi}{8}$, where μ and v are orientation and scale factors, respectively. By varying μ and v we can selected

different kernels. Fig. 2 shows an example for $\mu = 0, 1, \dots, 7$ and $v = 0, 1, \dots, 4$.

Given and 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).

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

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

$$F_{k,l} = \sum_{i=x_l-k}^{x_l+k} \sum_{j=y_l-k}^{y_l+k} |G_{ij}|, l = 0, 1, \dots, N; k = 1, 3, 5, 7, 9. \quad (5)$$

where N is the number of the fiducial points marked in the face image, x_l and y_l are the coordinates of the fiducial point l , and k is the number of neighboring pixels used to form the regions. Koutlas and Fotiadis (2008) proposed a set of 20 fiducial points which were derived from 74 different landmarks. 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. 3 shows the 20 fiducial points used in this work.

Here it is important to mention that those points can be defined either manually or automatically. In our research those points were manually located in the subjects face. For each fiducial point a mask of size $k \times k$ is used to compute the feature vector according to Eq. (5). In our experiments we have tested $k = \{1, 3, 5, 7, 9\}$.

As mentioned before, 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 classifiers.

3.2. Local binary patterns (LBP)

LBP operators have also been successfully applied to facial expression recognition (Shan et al., 2009) and for this reason they

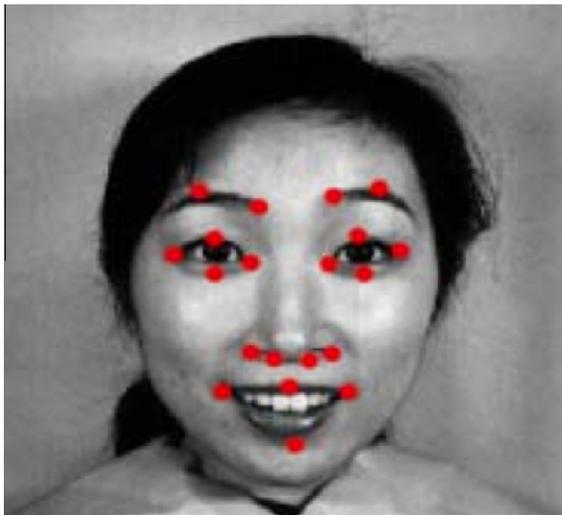


Fig. 3. The 20 fiducial points proposed by Koutlas and Fotiadis (2008).

were selected as another feature set used to train our base classifiers. The original LBP proposed by Ojala et al. (1996) labels the pixels of an image by thresholding a 3×3 neighborhood of each pixel with the center value and considering the results as a binary number and the 256-bin histogram of the LBP labels computed over a region is used as texture descriptor. Fig. 4 illustrates this process.

The limitation of the basic LBP operator is its small neighborhood which can not absorb the dominant features in large scale structures. To surpass this problem the operator was extended to cope with bigger neighborhoods (Ojala, Pietikinen, & Menp, 2002). Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. Fig. 5 exemplifies the extended LBP operator where (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.

The operator $LBP_{P,R}$ produces 2^P different output values corresponding to the 2^P different binary patterns that can be formed by the P pixels in the 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. An 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 (Ojala et al., 2002).

Accumulating the patterns which have more than two transitions into a single bin yields an LBP operator, denoted $LBP_{P,R}^{u2}$, 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 $LBP_{8,1}^{u2}$. Thereafter, an histogram of the frequency of the different labels produced by the LBP operator can be built.

According to Shan et al. (2009), an interesting way of using LBP in face images consists in equally divide the image into n small

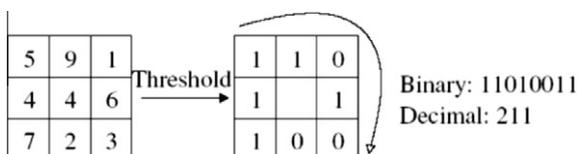


Fig. 4. The original LBP operator.

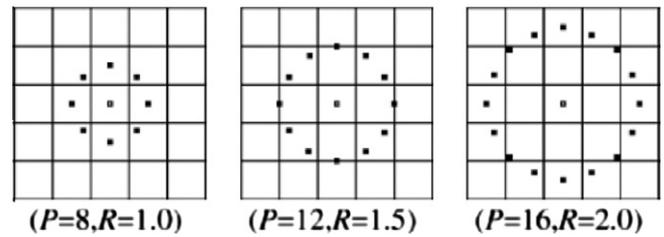


Fig. 5. Three examples of the extended LBP operator (Shan et al., 2009): the circular $(8;1)$ neighborhood, the circular $(12;1.5)$ neighborhood, and the circular $(16;2)$ neighborhood, respectively.

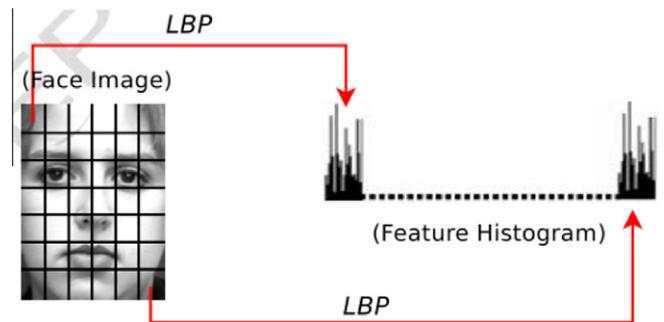


Fig. 6. LBP features extracted from a zoned face image (Shan et al., 2009).

zones Z_0, Z_1, \dots, Z_n to extract the LBP histograms. The features extracted from each zone are then concatenate into a single vector. Fig. 6 exemplifies this process.

In our experiments the faces were divided into 42 zones (7×6) . Three different configurations of the LBP operator were considered: $LBP_{8,1}^{u2}$, $LBP_{8,2}^{u2}$, $LBP_{16,2}^{u2}$. The first two configurations produce a feature vector of 59 components per zone, summing up 2478 components while the last one produces a feature vector of 243 components per zone, summing up, 10,206 components. In summary, we have 3 different feature configurations that will be used to train 3 classifiers.

4. Experiments and discussion

Two experimental protocols were employed to evaluate the proposed ensemble method for facial expression recognition. In Experiment I, subjects 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 the Experiment II, the subjects 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 experimental protocol is very often found in the literature. However, the second protocol is far more realistic since during the deployment phase the system would have to classify expressions from subjects that were not used to train the system.

We have employed SVMs as base classifiers. The computational cost of training a pool of SVMs is high, but on the other hand, the classification process is almost instantaneous because, given an instance, only its position in the feature space relative to the optimal hyperplane is evaluated. The pairwise strategy, where $d(d-1)/2$ classifiers are trained and organized as a tree, was employed due the facial expression recognition is a multi-class problem. Assuming that d denotes the number of classes we end up with twenty-one classifiers since we consider seven different classes of facial expressions. Such a tree is transposed from the leaves to the root, where the decision about the final class (facial expression) is taken.

Finally, it is important to mention that in all experiments we have used a 10-fold cross validation procedure, similar with that

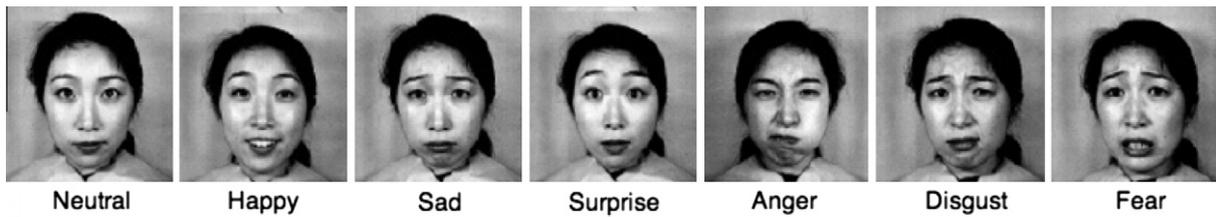


Fig. 7. Example of the seven categories of facial expressions taken from the JAFFE database.

used by Zhang et al. (1998). In the next subsections we describe the JAFFE and the Cohn-Kanade databases as well as we report the experiments carried out on both of these databases.

4.1. Databases

The JAFFE database (Lyons et al., 1998) contains 10 female subjects 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. An example of these categories is presented in Fig. 7. The names of the subjects are not revealed but they are referred with their initials: KA, KL, KM, KR, MK, NA, NM, TM, UY, and YM.

According to Bashyal and Venayagamoorthy (2008), each image in the database was rated by 91 experimental subjects for degree of each of the six basic expressions present in the image. The semantic rating of the images showed that the error for the fear expression was higher than that for any other expression but there exist a number of cases even for other expressions in which the expression getting highest semantic rating is different from the expression label of the image.

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 Cohn-Kanade database is encoded into combinations of action units. These combinations were translated into facial expressions according to Pantic and Rothkrantz (2000) in order to define the corresponding ground truth for the facial expressions. All the subjects were taken under consideration to form the database, composed of 1,281 images, for the experiments. Fig. 8 shows some examples of this dataset. Differently from the JAFFE Database where all the subjects have all the seven different facial expression, in this database few subject have the seven expressions. This leads to an unbalanced dataset.

4.2. Experiments on JAFFE database

According to the proposed methodology, the first step consists in training the pool of base classifiers. All the classifiers are SVMs trained with Gaussian kernel using LibSVM (Fan, Chen, & Lin, 2005). Kernel parameters such as C and γ were defined through a grid search using cross validation. Fig. 9 shows the accuracy of

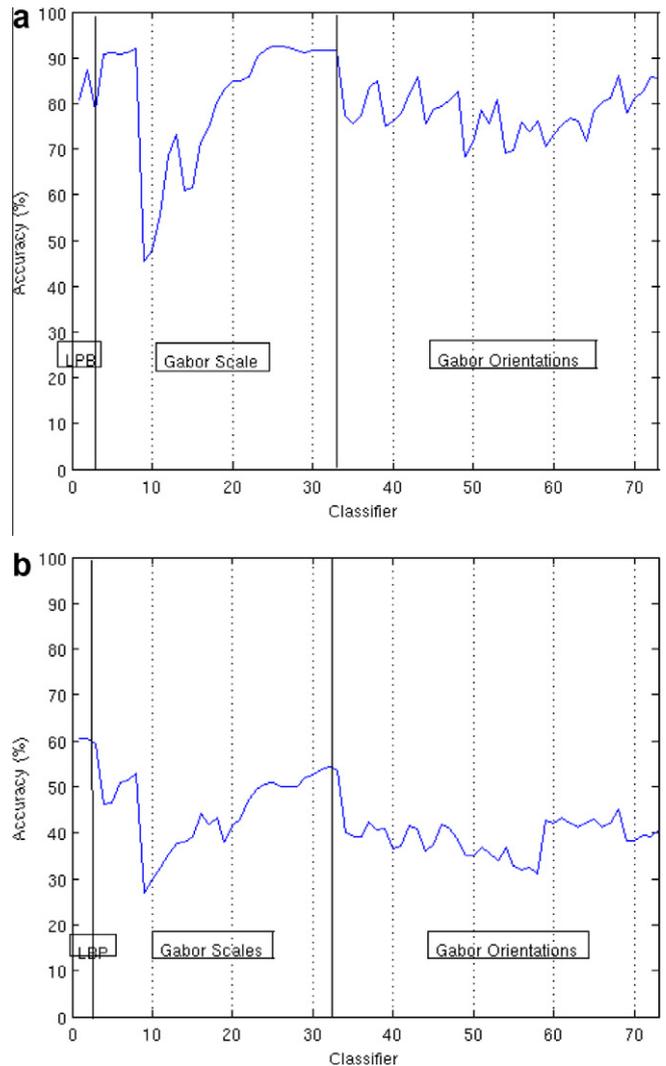


Fig. 9. Accuracy of the classifiers on JAFFE database: (a) Experiment I and (b) Experiment II.



Fig. 8. Example of the seven categories of facial expressions taken from the Cohn-Kanade database.

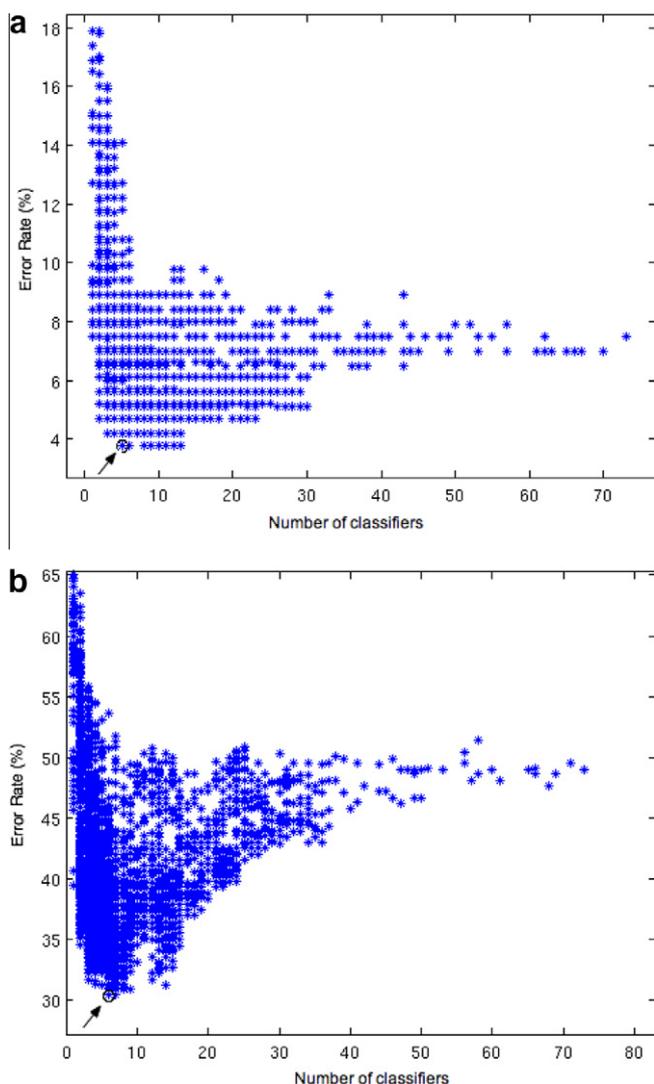


Fig. 10. Evolution of the population in the objective plane: (a) Experiment I and (b) Experiment II.

the 73 classifiers for experiments I and II using JAFFE database. The classifiers were split into three groups: 3 LBP, 30 Gabor scale-based, and 40 Gabor orientation-based classifiers. As we can observe, the performance of the classifier for the Experiment II is much worse than the performance achieved in Experiment I.

After training the pool of classifiers they are used as input to the MOGA. In this work we have used the NSGA II multi-objective genetic algorithm to build the ensemble of classifiers. The NSGA II 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 (Kittler et al., 1998). Other fusion rules such as Max, Min, Average, and Product were also tried out to compute the error rate of the ensemble but the Sum rule was the one that produced the best results. We have used the one-max problem to define the probabilities of crossover and mutation, since it is probably the most frequently used test function in research on genetic algorithms due to its simplicity (Cantu-Paz, 2000). The population size and the number of generations were empirically defined.

Fig. 10 shows the evolution of the population in the objective plane for Experiments I and II. As we can observe, in both cases the algorithm converges toward the Pareto-front producing a set of possible solutions. In order to perform the search here we also have used 10-fold cross validation. Each experiment was replicated 10 times to verify the reproducibility. Therefore, all the results presented here are the average of these 10 replications.

The next step consists in choosing the best ensemble of classifiers from the Pareto. As mentioned before, high accuracy is important but the size of the ensemble also is an important issue for this kind of application. As we can observe from Fig. 10 the ensembles that provide the best trade-off between accuracy and size are located close to the end of the Pareto. The selected ensemble are marked with an arrow in Fig. 10aa and b. The selected classifiers and their individual performances are reported in Table 1. Here it is important to remark that the selected ensembles were present in all the 10 replications, what guarantees that the ensembles were not found accidentally.

In spite of the same size (5 and 6 classifiers for Experiments I and II, respectively), the composition of the ensemble is totally different, with the exception of the LBP classifier LBP_{8,2}. As we can notice from Table 1 the problem of Experiment II is quite more difficult than the problem of Experiment I. However, the proposed methodology was able to find suitable ensembles for both experiments.

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 improve the recognition rate in about 10% relative to the best single classifier. 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 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%.

Tables 2 and 3 compare the confusion matrices for both experiments considering all the classifiers and the ensemble produced

Table 1
Selected classifiers – JAFFE database.

Experiment I		Experiment II	
Feature set	Accuracy (%)	Feature set	Accuracy (%)
LBP _{8,2}	87.3	LBP _{8,1}	60.6
Gabor scale 5, mask 3 × 3	91.6	LBP _{8,2}	60.6
Gabor orientation 3, mask 7 × 7	80.7	LBP _{16,2}	59.3
Gabor orientation 6, mask 7 × 7	76.6	Gabor orientation 2, mask 1 × 1	41.0
Gabor orientation 8, mask 7 × 7	85.9	Gabor orientation 3, mask 5 × 5	41.8
		Gabor orientation 6, mask 9 × 9	41.2
All classifiers	92.5	All classifiers	49.0
Ensemble	96.2	Ensemble	70.0

Table 2
Confusion matrices for Experiment I – JAFFE database.

	All classifiers							Ensemble						
	HA	FE	AN	SA	DI	SU	NE	HA	FE	AN	SA	DI	SU	NE
HA	28						3	28						3
FE		31				1			32					
AN			28		2					29		1		
SA	1	1		27	1		1	1	1		29			
DI			1	2	25	1				1		28		
SU			1			29							29	
NE						1	29							30

Table 3
Confusion matrices for Experiment II – JAFFE database.

	All classifiers							Ensemble						
	HA	FE	AN	SA	DI	SU	NE	HA	FE	AN	SA	DI	SU	NE
HA	17	4	1	3	2	2	2	27	3				1	
FE	1	15	2	3	7	1	3		20		3	5	4	
AN	2	4	17		7					24		5		1
SA	3	7	1	9	8		3	1	2	4	15	7	1	1
DI	2	5	4	1	16		1		1	3	5	20		
SU	1	3	1		1	29	2	1	3				24	2
NE	4	4		2	1	6	13	1	4	3			3	19

Table 4
Comparison with different approaches on JAFFE database.

Reference	Accuracy (%)	Features
Zhang et al. (1998)	90.1	Geometry and Gabor
Bashyal and Venayagamoorthy (2008)	90.2	Gabor and LVQ
Koutlas and Fotiadis (2008)	92.3	Gabor filters
Liu and Wang (2006)	92.5	Gabor filters
Oliveira et al. (2011)	94.0	2DPCA with feature selection and SVM
Shih et al. (2008)	94.1	2D-LDA and SVM
Liao et al. (2006)	94.5	LPB, Tsallis entropies, global appearance
Cheng et al. (2010)	95.2	Gaussian process
Zhi and Ruan (2008)	95.9	2D locality preserving projections
Proposed approach	96.2	Ensemble based on Gabor and LBP

by the proposed method. Table 2 shows us that most confusions of the Experiment I have been solved by the ensemble. In Experiment II, several confusions also have been solved, e.g., class Sad (SA), however, there is a lot of room for improvement. A possible alternative to further reduce these confusion would be to use images from other databases to increase the training set.

Table 4 shows the performance of different approaches reported in the literature on JAFFE database. To the best of our knowledge, all works have used the protocol we have applied 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, we can see that the proposed methodology compares favorably to the literature.

4.3. Experiments on Cohn-Kanade database

The same protocol used for JAFFE database was applied to the experiment on Cohn-Kanade. Fig. 11 shows the accuracy of the 73 classifiers for experiments I and II using Cohn-Kanade database. Here the classifiers were separated into the same three groups and the like in the previous experiment the performance of the classifier for the Experiment II is worse than the performance achieved in Experiment I.

However, by comparing Figs. 9 and 11 it is clear that the Cohn-Kanade database is less complex than the JAFFE database. This could be explained by the fact that facial expression images were extracted from video sequences which reduces considerably the variability of the same subject, as depicted in Fig. 12. This explains the compelling performance of some classifiers, especially in Experiment I where the same subject participate in both training and testing sets.

Alike the experiments on JAFFE database, here the algorithm also converges toward the Pareto-front producing a set of possible solutions. The selected ensemble are marked with an arrow in Fig. 13a and b. The selected classifiers and their individual performances are reported in Table 5. Again, the selected ensembles were present in all the 10 replications what guarantee that they were not found accidentally.

As mentioned before, this dataset is less complex than the previous one so 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

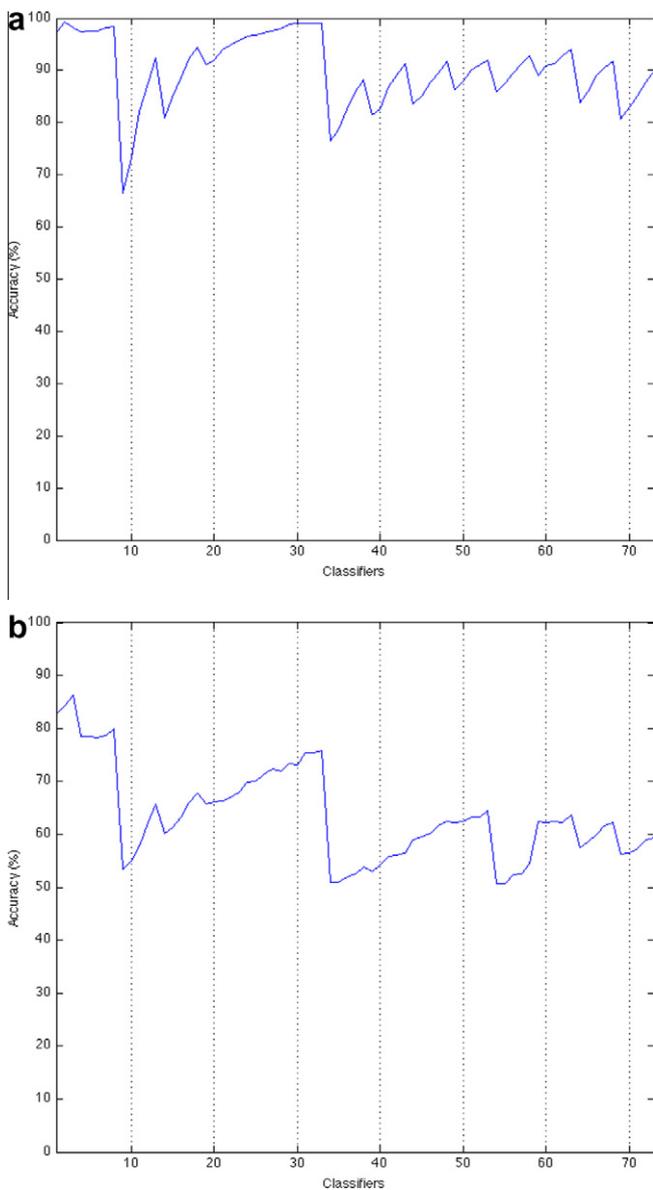


Fig. 11. Accuracy of the classifiers on Cohn-Kanade database: (a) Experiment I and (b) Experiment II.



Fig. 12. Small variability of the Cohn-Kanade database.

classifiers can bring important information to the ensemble. Table 6 shows the confusion matrix for Experiment II where we can observe that several confusions with the class “Fear” were solved. According to Zhang et al. (1998), fear is the most difficult expression to be recognized, even by humans.

Table 7 shows the performance of different approaches reported in the literature on Cohn-Kanade database. However, a direct

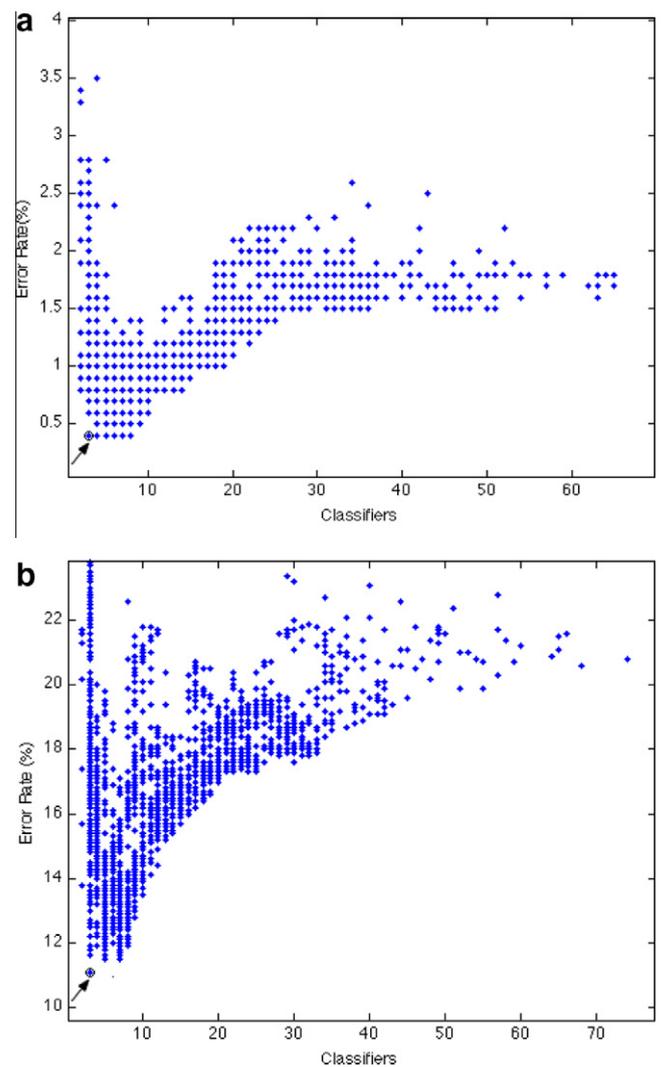


Fig. 13. Evolution of the population in the objective plane: (a) Experiment I and (b) Experiment II.

comparison is not possible due to the differences in the experimental protocol. For instance, Shan et al. (2009) and Bartlett, Littlewort, Fasel, and Movellan (2003) have partitioned the dataset randomly into groups of roughly equal numbers of subjects where one group was used as the test data, while the remaining groups were used as the training data to train classifiers. We can see that the proposed methodology compares favorably to the literature regardless the differences in the experimental protocol.

5. Conclusion

In this paper, we have presented a novel method for facial expression recognition that relies on the combination of two different feature sets in an ensemble approach to improve the recognition accuracy. The proposed approach combines two different features sets, namely Gabor filters and LBP that operate in different representation spaces. The recognition rate resulting from the combination of both feature sets into an ensemble of classifiers is significantly better than that achieved by individual features sets and single classifiers. For instance, in the case of Experiment I, the ensemble brought an improvement of about 5% compared to the best individual classifier. A more impressive improvement was achieved in Experiment II where the ensemble improves the

Table 5
Selected classifiers – Cohn-Kanade database.

Experiment I		Experiment II	
Feature set	Accuracy (%)	Feature set	Accuracy (%)
LBP _{8,2}	99.0	LBP _{8,2}	84.3
Gabor scale 6, mask 1 × 1	98.7	Gabor scale 1, mask 7 × 7	78.7
All classifiers	98.3	All classifiers	79.2
Ensemble	99.4	Ensemble	88.9

Table 6
Confusion matrices for Experiment II – Cohn-Kanade database.

	All classifiers							Ensemble						
	HA	FE	AN	SA	DI	SU	NE	HA	FE	AN	SA	DI	SU	NE
HA	234	10			1	1	15	247	3					11
FE	18	70	3	6	8	19	29	12	122		1		3	15
AN	4	8	37	8	9	3	15	5	1	45	8	12		13
SA		5	7	105	2	7	27		2	11	121		1	18
DI	11	3	8	8	74	4	6			15	5	86		8
SU	5			2		212	6	3	1				218	3
NE	9	5	5	2		4	270	5	2	3	1			280

Table 7
Comparison with different approaches on Cohn-Kanade database.

Reference	Accuracy (%)	Features
Shan et al. (2009)	79.1	LBP + template matching
Cohen et al. (2003)	73.2	Geometric + Tree-augmented-Naive Bayes
Bartlett et al. (2003)	86.9	Gabor filter + SVM
Bartlett et al. (2005) (Exp. II)	89.1	Gabor filter + SVM
Shan et al. (2009)	88.9	LBP-based + SVM
Shan et al. (2009)	86.8	Gabor filter + SVM
Proposed approach (Exp. I)	99.4	Ensemble based on Gabor and LBP
Proposed approach (Exp. II)	88.9	Ensemble based on Gabor and LBP

recognition rate in about 10% compared with the best individual classifiers.

Compared with other results available in the literature that use the same experimental protocol (Bashyal & Venayagamoorthy, 2008; Cheng, Yu, & Xiong, 2010; Koutlas & Fotiadis, 2008; Liao, Fan, Chung, & Yeung, 2006; Liu & Wang, 2006; Oliveira, Koerich, Mansano, & Britto, 2011; Shih, Chuang, & Wang, 2008; Zhang et al., 1998; Zhi & Ruan, 2008), the results reported in this paper represent a slight improvement in terms of recognition rate. Recent works in facial expression recognition report recognition rates between 90% and 96%. It is important to notice that the two databases do not convey realistic scenario regarding the acquisition of samples. Situations such as low and changing illumination, noise addition or scaling are not addressed in both databases. However, such databases are publicly available and have been used by many researchers for evaluation and benchmarking.

In spite of the good results achieved, there are some shortcomings related to the proposed approach. The first shortcoming is the necessity of locating the fiducial points in the case of the Gabor features. Since there is no reliable algorithm to locate such points in a face image, the incorrect location leads to noisy feature vectors which can decrease the accuracy of the corresponding classifier. However, in the scope of this paper, it would be impractical to study the impact of the mislocation of fiducial points for the ensemble. Nevertheless, this problem can be somehow alleviated by the ensemble. As stated in Section 4, even if a classifier presents a poor performance it could be important to the ensemble. Another shortcoming is the increase of the complexity of the whole system since it requires the extraction of two sets of features and the

training and selection of the classifiers. Since this additional computational effort is only required at the developing phase, the percent rise in the facial expression recognition rate afforded by the proposed approach is worthwhile.

In summary, the main contribution of this paper is a novel approach that creates ensemble of classifiers from a pool of base classifiers trained with two feature sets which are widely used for automatic facial expression recognition. By varying the parameters of the Gabor filters and LBP, seventy-three classifiers were trained and further used as input of a multi-objective genetic algorithm that returns a set of possible ensembles. The proposed approach is effective and the improvements reported in this paper are significant. Hence, it is logical to conclude that ensemble of classifiers is a promising research direction in facial expression recognition.

Acknowledgments

The authors would like to acknowledge the National Council for Scientific and Technological Development (CNPq) for the financial support under the Grants 471.496/20071-3, 309.295/2007-6 and 306.703/2010-6.

References

- Aleksic, P. S., & Katsaggelos, A. K. (2006). Automatic facial expression recognition using facial animation parameters and multistream hmms. *IEEE Transactions on Information Forensics and Security*, 1(1), 3–11.
- Bartlett, M., Littlewort, G., Fasel, I., & Movellan, R. (2003). Real time face detection and facial expression recognition: Development and application to human computer interaction. In *Computer Vision and Pattern Recognition Workshop* (p. 53).

- Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005). Recognizing facial expression: Machine learning and application to spontaneous behavior. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 568–573).
- Bashyal, S., & Venayagamoorthy, G. K. (2008). Recognition of facial expressions using gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence*, 28, 1056–1064.
- Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 711–720.
- Besinger, A., Szynda, T., Lal, S., Duthoit, C., Agbinya, J., Jap, B., et al. (2010). Optical flow based analyses to detect emotion from human facial image data. *Expert Systems with Applications*, 37, 8897–8902.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Cantu-Paz, E. (2000). *Efficient and accurate parallel genetic algorithms*. Kluwer Academic Publishers.
- Cheng, F., Yu, J., & Xiong, H. (2010). Facial expression recognition in jaffe dataset based on gaussian process classification. *IEEE Transactions on Neural Networks*, 21(10), 1685–1690.
- Cohen, I., Sebe, N., Garg, A., Chen, L., & Huang, T. S. (2003). Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91, 160–187.
- Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. John Wiley and Sons Ltd.
- Deb, K., Agarwal, A., & Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: Nsga-ii. *IEEE Transaction on Evolutionary Computation*, 6(2), 181–197.
- Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality Social Psychology*, 17(2), 124–129.
- Fan, R.-E., Chen, P.-H., & Lin, C.-J. (2005). Working set selection using the second order information for training svm. *Journal of Machine Learning Research*, 6, 1889–1918.
- Fasel, B., & Luettnb, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition*, 36, 259–275.
- Freund, Y., & Schapire, R. (1996). Experiments with a new boosting algorithm. In *Proceedings of 13th international conference on machine learning* (pp. 148–156). Bary, Italy.
- Geetha, A., Ramalingam, V., Palanivel, S., & Palaniappan, B. (2009). Facial expression recognition – A real time approach. *Expert Systems with Applications*, 36, 303–308.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- Kittler, J., Hatef, M., Duin, R. P., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 226–239.
- Koutlas, A., & Fotiadis, D. I. (2008). An automatic region based methodology for facial expression recognition. In *IEEE International Conference on Systems, Man and Cybernetics* (pp. 662–666).
- Kuncheva, L. (2004). *Combining pattern classifiers. Methods and algorithms*. New York: Wiley.
- Liao, S., Fan, W., Chung, C. S., & Yeung, D.-Y. (2006). Facial expression recognition using advanced local binary patterns. In *International conference on image processing (ICIP)* (pp. 665–668). IEEE.
- Liu, W., & Wang, Z. (2006). Facial expression recognition based on fusion of multiple gabor features. In *18th International conference on pattern recognition 2006* (pp. 536–539).
- Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12), 1357–1362.
- Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Third IEEE international conference on automatic face and gesture recognition* (pp. 200–205).
- Mase, K., & Pentland, A. (1991). Recognition of facial expression from optical flow. *IEICE Transactions*, 74(10), 3474–3483.
- Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29(1), 51–59.
- Ojala, T., Pietikainen, M., & Menp, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987.
- Oliveira, L. E. S., Koerich, A. L., Mansano, M., & Britto, A. S. Jr., (2011). 2d principal component analysis for face and facial-expression recognition. *Computing in Science and Engineering*, 13(3), 9–13.
- Oza, N. C., & Tumer, K. (2008). Classifier ensembles: Select real-world applications. *Information Fusion*, 9(1), 4–20.
- Pantic, M., & Patras, I. (2006). Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics*, 36(2), 433–449.
- Pantic, M., & Rothkrantz, L. J. M. (2000). Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 18(11), 881–905.
- Santos, E. M., Sabourin, R., & Maupin, P. (2009). Overfitting-cautious selection of ensembles of classifiers with genetic algorithms based on complexity, diversity and accuracy criteria. *Information Fusion*, 10, 150–162.
- Shan, C., Gong, S., & McOwan, P. (2005). Robust facial expression recognition using local binary patterns. In *IEEE International Conference on Image Processing* (pp. 370–373).
- Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27, 803–816.
- Shih, F. Y., Chuang, C.-F., & Wang, P. S. P. (2008). Performance comparisons of facial expression recognition in JAFFE database. *International Journal of Pattern Recognition and Artificial Intelligence*, 22(3), 445–459.
- Tan, X., & Triggs, B. (2007). Fusing gabor and lbp feature sets for kernel-based face recognition. In *Proceedings of the third international workshop analysis and modelling of faces and gestures* (pp. 235–249). Rio de Janeiro, Brazil.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1, 71–86.
- Wong, J. J., & Cho, S. Y. (2009). A local experts organization model with application to face emotion recognition. *Expert Systems with Applications*, 36, 804–819.
- Xiea, S., Shana, S., Chena, X., Mengc, X., & Gao, W. (2009). Learned local gabor patterns for face representation and recognition. *Signal Processing*, 89(12), 2333–2344.
- Zavaschi, T., Oliveira, L., & Koerich, A. (2011). Facial expression recognition using ensemble of classifiers. In *Proceedings of 36th international conference on acoustics, speech and signal processing* (pp. 1489–1492).
- Zhang, Z., Lyons, M. J., Schuster, M., & Akamatsu, S. (1998). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Third IEEE international conference on automatic face and gesture recognition* (pp. 454–459).
- Zhang, Y., & Ji, Q. (2005). Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5), 1–16.
- Zhi, R., & Ruan, Q. (2008). Facial expression recognition based on two-dimensional discriminant locality preserving projections. *Neurocomputing*, 71, 1730–1734.