Facial Expression Recognition Using a Pairwise Feature Selection and Classification Approach

Marcelo J. Cossetin, Julio C. Nievola
Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná
Curitiba, PR, Brazil 80215-901
Email: nievola@ppgia.pucpr.br

Alessandro L. Koerich
Department of Software and IT Engineering
École de Technologie Supérieure
Montréal, QC, Canada H3C 1K3
Email: alessandro.koerich@etsmtl.ca

***Abstract*—This paper proposes a novel approach that combines specialized pairwise classifiers trained with different feature subsets for facial expression classification. The proposed approach first detects and extracts automatically faces from images. Next, the face is split into several regular zones and textural features are extracted from each zone to capture local information. The features extracted from all zones are concatenated to model the whole face. A pairwise approach that considers all pairs of classes and a hybrid feature selection strategy is used to both reduce the dimensionality and to select relevant features to discriminate between specific pairs of classes. Several pairwise classifiers are then trained with such pairwise feature subsets. At the end, given a new face image, all features are extracted from such a face, but only the previously selected subset of features is inputted to each pairwise classifier. The output of all pairwise classifiers is combined using a majority voting rule to decide on the facial expression. Experiments have been carried out on three publicly available datasets (JAFFE, CK and TFEID) and the correct classification rates of 99.05%, 98.07% and 99.63% were achieved respectively. Therefore, the pairwise approach is effective to discriminate between different facial expressions and the results achieved by the proposed approach are slightly better than several current approaches.**

I. Introduction

With the recent advances in computer vision, the analysis of facial expression is being used more often in many tasks. The recognition of facial expressions is an active research subject for more than two decades [1], [2]. In a conversation between humans, the facial expressions form a communication channel that brings important information about the mental, emotional and physical states of an individual [3]. The study conducted by Mehrabian [4], pointed out that the human communication is 7% verbal, 38% vocal and 55% by facial expressions. Furthermore, Ekman and Friesen [5] claim that there are cross-cultural universal facial expressions that belong to six groups: happiness (HA), sadness (SA), disgust (DI), fear (FE), anger (AN) and surprise (SU). There is also the neutral (NE) category.

An automatic system for facial expression recognition is made up of face detection, feature extraction, dimensionality reduction and classification [6]. The face detection within an image has been carried out using the Viola-Jones [7] algorithm since it is a fast and efficient method to detect objects. This step is very important because it allows the segmentation of the area of interest where further processing will be applied.

Recent works have shown that geometrical approaches that rely on fiducial points are the most successful for facial expression recognition and other related problems [8]–[12]. The geometrical approach with fiducial points usually leads to high accuracy but it depends on the precise location of the face elements which can be flawed by the variations imposed during the image capture such as illumination, pose, etc. Furthermore, the extraction algorithms produce a high number of features since usually a feature vector is extracted from each of such fiducial points or from zones delimited by such points. The principal component analysis (PCA) algorithm has been used to reduce the dimensionality of the feature vectors [12], [13], but PCA is a global technique that evaluates the whole dataset without relying into the classes.

An alternative to the geometric methods are the appearance-based methods such as Local Binary Pattern (LBP) and Gabor Filters, that do not require the explicit location of fiducial points. Several local features have been proposed to extract micro patterns from facial zones such as the Median Ternary Pattern (MTP) [8], Weber Local Descriptor [3], and the popular LBP and its variations such as Monogenic Binary Patterns and Local Ternary Patterns. Usually the face image is split into several zones and feature vectors are generated for each zone. The final feature vector is made up by the concatenation of the feature vectors from each zone and this may result in a high dimensional vector. For instance, Bashar et al. [8] have used the MTP which produces 512 features for each of the 7×6 face zones and the final feature vector has 21,504 dimensions.

The high dimensionality present in most of the current methods for facial expression recognition may affect the performance of the classifiers due to the possible presence of redundant features that do not contribute to discriminate between classes and increase the computational effort. Therefore, feature selection and dimensionality reduction techniques have been used to found subset of features for facial expression recognition. Hussain et al. [3] have used dimensionality reduction techniques such as PCA and Kruskall Wallis (KW) to select the most discriminant features; however, they did not have evaluated the impact of such a reduction neither in accuracy nor in the computational time.

Support Vector Machines (SVM) has been the main machine

learning algorithm used in facial expression recognition. In Verma et al. [10], the AdaBoost algorithm with weak classifiers has achieved a higher accuracy than SVM, however with a higher computational cost. Martin et al. [14] have compared a multilayer Perceptron (MLP) and SVM and the accuracy was 75% and 92%, respectively. In Liu et al. [9] a SVM has also achieved a better recognition rate than a k-Nearest Neighbor (KNN) classifier. Hussain et al. [3] have used ensemble of classifiers for facial expression recognition, where the classifiers were specialized in a single facial expression. Kyperountas et al. [15] have used the dimensionality reduction based on pairs of facial expression where subsets of features are selected to achieve a better discrimination between pairs of facial expressions and classification one-against-one. For the JAFFE dataset, a recognition rate of 95.11% was achieved. In spite of the promising results, the Gabor filter is used as feature extractor which requires the knowledge of the fiducial points and also demands a high computational effort making difficult its use in real-time applications [8]. Furthermore, the face detection was done manually and the validation of the approach was carried out on a single dataset. The number of features that has been selected is not presented in the paper.

This paper proposes a novel approach for facial expression recognition that leads to high-discriminant, low-dimensionality feature vectors made up of different feature subsets. The proposed approach splits a multi-class problem into several binary problems at both feature and classification levels. At the feature level, feature selection is carried out by a hybrid method that considers each possible pairs of classes, and select the features that have the highest discriminate power for each pair. Therefore, we end up with class-dependent feature subsets which are used to train class-dependent binary classifiers. At the classification level, given an input face image, all features are extracted but the binary classifiers use only the corresponding subsets. At the end, the outputs of all pairwise classifiers are combined using a simple majority voting rule. The proposed approach achieves a high accuracy demanding less computation power, which is very competitive with the current state-of-the art.

This paper is organized as follows: Section II presents the details of the proposed approach to detect faces, extract features, reduce the dimensionality and classification. The experimental results on three public datasets Japanese Female Facial Expression (JAFFE), Cohn-Kanade (CK) and Taiwanese Facial Expression Image Dataset (TFEID) are presented in Section III. The conclusions are stated in the last section.

II. Proposed Approach

In this paper we propose a pairwise feature selection and classification approach that takes, for each pair of classes, the best discriminatory features and use them to train a specialized classifier. Therefore, the feature subsets may be different, depending on the pair of classes. On the other hand, the conventional approaches consider all classes together and a single subset of features is selected to represent the whole problem [16] and used to train a multi-class classifier to distinguish among all classes. Different from the conventional use of PCA for feature selection [17] where the features are transformed to a new space without relying on the class discrimination, the proposed approach retains the original features and just selects a subset of them. Since the feature selection process considers pairs of classes, the classification is carried out using an ensemble of binary classifiers in a one-against-one strategy and the final decision is taken using the majority vote rule.

Fig. 1 shows an overview of the proposed approach that includes facial detection, feature extraction, pairwise feature selection and reduction and classification which are described as follows.

A. Face Detection

The first step is to locate and segment a face within an image to minimize the impact of other image elements as well as to normalize the face images. Faces are located using the algorithm proposed by Viola and Jones [7] and has been used in face recognition due to its accuracy and low computation cost [10] when compared to other techniques. The same algorithm is used to detect the eyes within the detected face.

Once a face is located, it is segmented and undergoes an histogram equalization to enhance the contrast as shown in Fig. 2a. All the images where faces are not correctly segmented are discarded at this step to improve the robustness of the proposed approach.

B. Feature Extraction

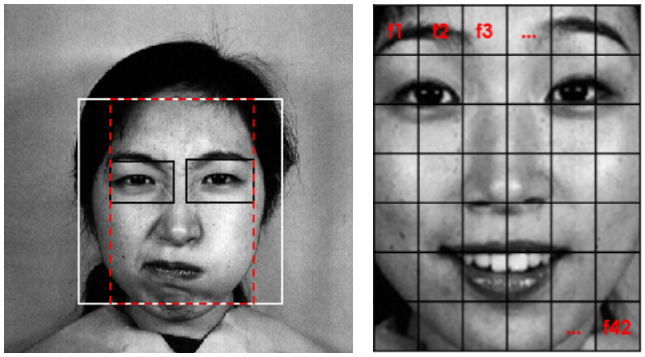
After the face detection step, features are extracted from the segmented images. The face images are modeled by two textural features: local binary patterns (LBP) [11] and Weber local descriptor (WLD) [6]. Since we are dealing with different datasets, after the segmentation the face images were normalized to the average dimension of 134×114 pixels as suggested in [6] and [9]. Since we want to capture local information related to particular regions of the faces and extract micro patterns, we use a zoning schema to split the faces images, as shown in Fig. 2b. At the end, the features extracted from each region are concatenated into a single feature vector that represents the face image.

For the $LBP^{u2}_{(8,2)}$ operator, the face image is split into 6×7 zones [8] and into 6×9 zones for the WLD extractor with $T = 8$ and $C = 5$ [9]. It is difficult to determine the number of zones since such a number must be a trade-off between the face representation and the computation time. This implies in different splitting schemes for comparison [18]. As the parameters of the feature extraction algorithm, the face's number of divisions should be set based on the size of the images being used. Therefore, different scenarios require different configurations.

C. Pairwise Feature Selection

The zoning approach employed during the feature extraction step leads to a high dimensional feature vector. Feature selection techniques are then used to select relevant features

Fig. 1. Overview of the proposed approach for facial expression recognition



(a) (b)

Fig. 2. Face detection and zoning: (a) face location (white), eyes detection (black) and face region delimitation (red) within an image; (b) zoning of the segmented face by a 6×7 grid into 42 zones ($f_1, f_2, \ldots, f_{42}$).

and reduce the dimensionality of the feature vector. This work proposes a pairwise feature selection approach that looks for subsets of features that best discriminate pairs of facial expressions where it is carried out an exhaustive evaluation of all possible pairs of facial expressions $p \in \{AN \cup SU, AN \cup HA, \ldots, DI \cup NE\}$.

Therefore, $C(C-1)/2$ different feature subsets are selected from F'_p, where C is the number of facial expressions, each corresponding to a pair of facial expressions p. Each feature subset is used to train a binary classifier as shown in Fig. 3b.

To find the best subset F'_p that separates the expressions of p, two approaches were used: a filter and a hybrid approach. In the hybrid approach a filter and a wrapper method are combined. First, a filter (eg CFS, ReliefF, IG, Mutual Information) is used to rank the attributes according to their relevance to separate the facial expressions and a condition is set to hold the most important attributes [19]. The wrapper uses these attributes to evaluate the classification rate based on a specific learning algorithm.

Unlike the literature, in our hybrid approach, the wrapper method does not evaluate all subsets of features, but it starts with a subset containing a predetermined small number of the most promising features, and the most promising features are added in the subset to evaluate the classification rate. The features selected by the wrapper method are used in the classification stage.

D. Classification

The main characteristic of the proposed approach is that we end up with several pairwise classifiers. Therefore, to recognize a facial expression we use a one-against-one classification approach. Since we have seven different facial expressions, including the neutral, twenty-one classifiers are used, each one aiming to distinguish between a pair of facial expressions. The novelty of the proposed approach is that each pairwise classifier c_p uses a particular subset of the whole feature set (F) as shown in Fig. 4c. Such 21 classifiers are combined used a simple majority vote rule and the facial expression that has the highest number of votes is chosen.

The results obtained with the pairwise approach (4b) are compared with a conventional approach that uses a single multi-class classifier (4a). This step is carried out to verify the number of attributes selected by each approach and to compare the computational cost.

III. Experimental Results

The experiments to evaluate the proposed approach were carried out on three datasets: Japanese Female Facial Expression (JAFFE) which is made up by 10 subjects with 3 to 4 images of each of the seven standard facial expression summing up to 213 images; Cohn-Kanade which comprises a sequence of images from neutral to a given expression of 123 subjects between 18 and 30 years old summing up to 427 images; Taiwanese Facial Expression Image Dataset which is made up of 40 subjects, each with 8 expressions summing up 268 images. For the Cohn-Kanade dataset only the last images of the sequences are used in the experiments and the neutral faces are taken from the first images. The detection of faces within the images is not used with the TFEID dataset because the images already have well-segmented faces.

In the literature there is a difference in the validation method. Several studies use cross-validation [8], [11], [20], while others separate datasets in training and testing subsets [21]. There are also studies that use only a small number of faces to evaluate the method [22]. Thus, the validation method can produce a high hit rate to less robust methods.

For dimensionality reduction, the correlation-based feature selection (CFS) is used as filter method with a stop criterion of five nodes that does not improve the merit of the set [23]. The CFS is a standard method for feature selection and it is able to select features that are highly correlated to a class and it is not commonly used in facial expression recognition [16].

The hybrid dimensionality reduction employs the information gain (IG) which has been used in facial expression recognition [24] and the Kruskal-Wallis (KW) method that has also been used in facial expression recognition [3] and additionally has a low computational cost. Since both methods only rank the features, following the ranking, a range between the 20-best and the 300-best first features was evaluated using a classification algorithm and a 10-fold cross-validation procedure [16].

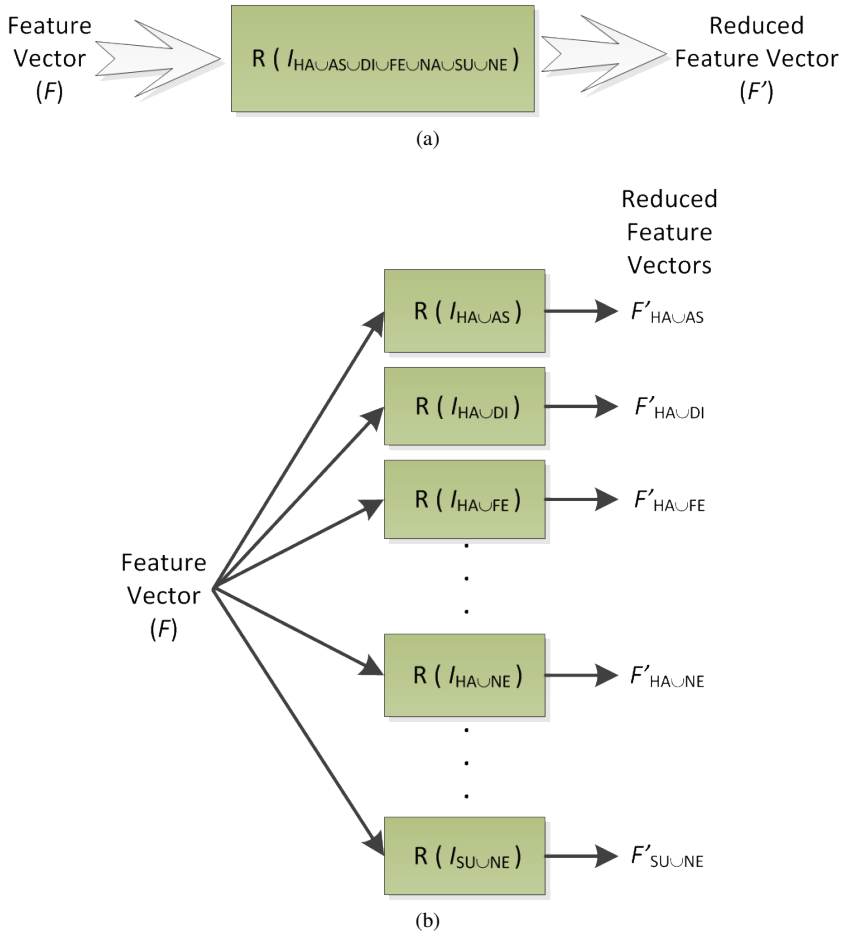


Fig. 3. Approaches for feature selection and dimensionality reduction: (a) global feature selection strategy that considers all expressions and results into a single feature subset (F'); (b) pairwise feature selection strategy that considers pairs of expressions and results in multiple feature subsets ($F'_{HA\cup AS}, \cdots, F'_{SU\cup NE}$).

Support vector machines with a Gaussian kernel were used as pairwise classifiers. A grid search was employed to find the value of the best parameters ($2^{-5} \leq C \leq 2^{15}$ and $2^{-15} \leq \gamma \leq 2^{3}$). We have also evaluated the k Nearest Neighbours algorithm but the results are not reported in this paper.

Tab. I shows that 2 and 12 faces were not detected for the JAFFE and CK datasets respectively. The misdetection is mainly due to the presence of hair close to the eyes as well as the state of the eyes (close or semi-closed) as shown in Fig. 5. Therefore, in the next steps we use 211 and 415 images from JAFFE and CK datasets respectively

TABLE I
RATE OF CORRECTED DETECTED FACES

Dataset	# of images	# detected faces	detection rate (%)
JAFFE	213	211	99
CK	427	415	97

Tab. II shows the number of features selected by each of the dimensionality reduction approaches, considering a multi-class problem where the feature vectors of all classes are used together, and the pairwise approach where the dimensionality reduction algorithms are applied only on the feature vectors of pairs of classes. Since we consider seven facial expressions, we have twenty one subsets of features that are submitted to the dimensionality reduction algorithm. In this case, the values shown in Tab. II for pairwise represent the average value for the twenty one subsets.

In average, we have 199 features for the global feature selection approach while in the pairwise feature selection approach this number drops to 72. This reduction is statistically significant according to the non-normality Shapiro-Wilk test followed by the Mann-Whitney ($\alpha = 0.5$) non-parametric test. Tab. III shows the results of the facial expression classification using both the global (global FS) and the pairwise (pairwise FS) feature selection approach as well as using the complete feature set (full). The best classification rates were achieved using the pairwise FS approach: 99.05% using LBP features or WLD features with the hybrid method (KW+SVM) for the JAFFE dataset; 98.07% using LBP features with the hybrid method (IG+SVM) for the Cohn-Kanade dataset; 99.63% using LBP features with the hybrid method (KW+SVM) for the TFEID dataset. Furthermore, the differences between the pairwise FS approach and the global FS approach shown

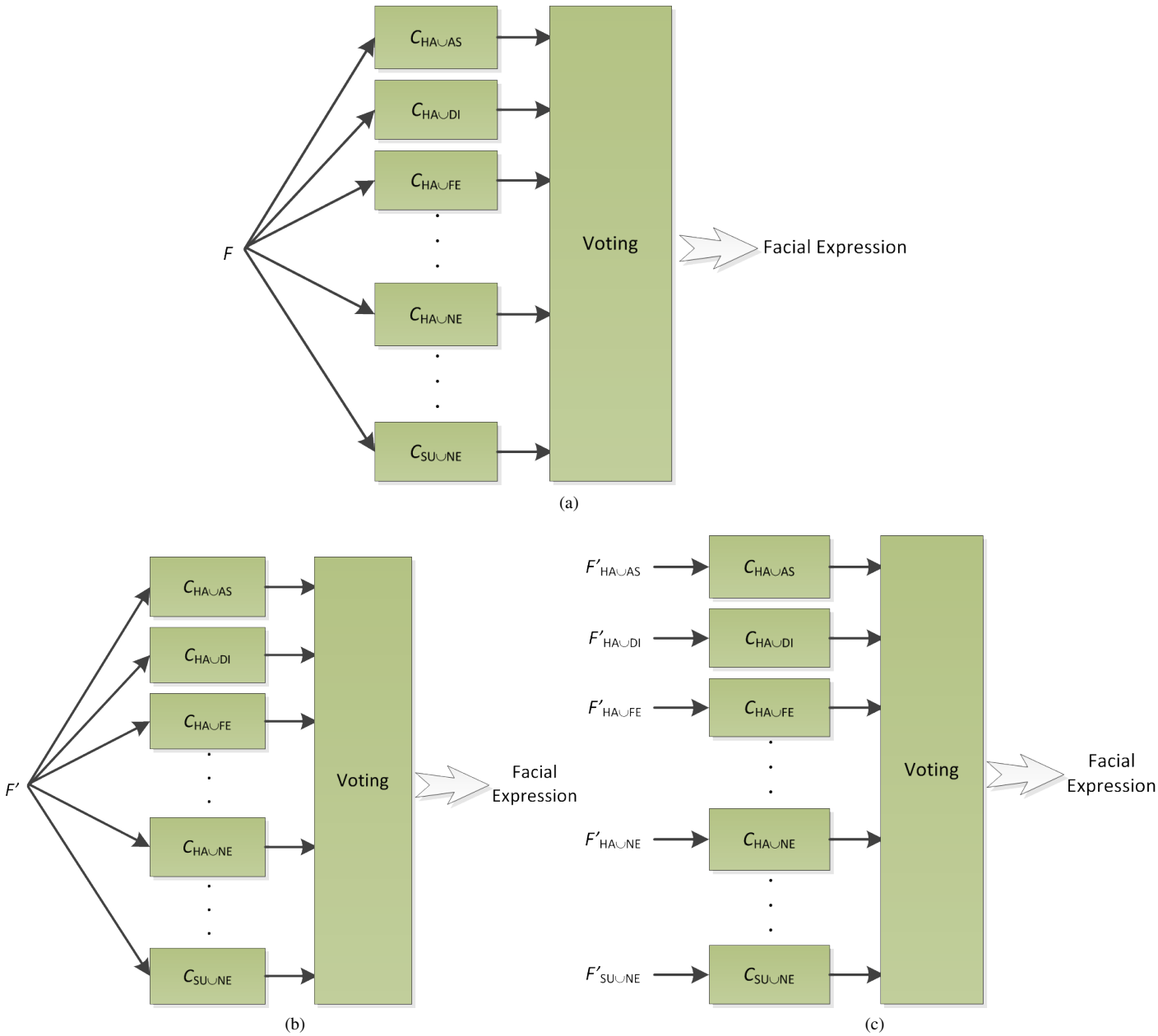


Fig. 4. Approaches for facial expression classification based on pairwise classifiers trained with global features (F and F') and pairwise features ($F'_{(.)}$): (a) conventional pairwise classification approach where all classifiers use the same feature set (F); (b) conventional pairwise classification approach where all classifiers use the same feature subset (F'); (c) proposed approach where each pairwise classifier uses a different subset of selected features ($F'_{(.)}$)

TABLE II
NUMBER OF SELECTED FEATURES IN EACH DATASET FOR THE GLOBAL AND PAIRWISE FEATURE SELECTION APPROACHES.

Feature Set	Approach	JAFFE		CK		TFEID	
		global FS	pairwise FS	global FS	pairwise FS	global FS	pairwise FS
WLD	filter CFS	58	60	145	101	135	81
WLD	hybrid IG+SVM	80	72	300	83	120	53
WLD	hybrid KW+SVM	280	74	260	77	300	78
LBP	filter CFS	64	59	137	91	136	78
LBP	hybrid IG+SVM	120	43	280	108	280	34
LBP	hybrid KW+SVM	280	48	220	98	260	33

(a) (b)

Fig. 5. Samples of misdetected faces in: (a) JAFFE dataset and (b) Cohn-Kanade (CK) dataset.

in Fig. 6 are statistically significant according to the Mann-Whitney ($\alpha = 0.05$) test.

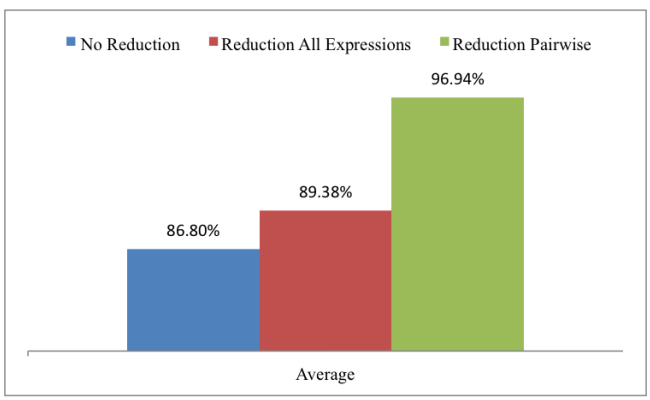


Fig. 6. Comparison of the correct classification rates without feature selection, with global feature section and with the proposed pairwise feature selection.

Tab. IV shows the results achieved by the proposed approach compared to other recent approaches. In spite of providing higher recognition rates than the current approaches in most of the cases, such a difference is not statistically significant in some cases. For the JAFFE dataset, the difference is significant to Wang et al. [6] but no difference was found to Zavaschi et al. [11]. For the CK dataset, Bashar et al. [8] has shown a statistically significant inferior performance while there is no significant difference to Wang et al. [6]. The statistical tests were not used to compare the results against the rest of the works due to the lack of information. Therefore, the proposed approach achieves a performance similar or higher than the works reported recently in the literature.

Besides the correct recognition rate, we have also compared the computational efforts of the proposed approach with other works. For such an aim, we have considered the number of features used in each classifier as metrics, which is similar to the metric proposed by Last et al. [25]. Bashar et al. [8] use feature vector of 21,504 dimensions with seven one-against-all SVMs, therefore $21,504 \times 7 = 150,528$ features are processed. The same happens in Zavaschi et al. [11] and Liu et al. [9] [6]. Even using 21 classifiers, the pairwise features selection strategy is able to select a small number of features for each SVM relative to the other approaches while providing a superior or equivalent performance.

IV. Conclusion

We have proposed a new approach to build specialized pairwise classifiers where features are selected according to the pair of classes. Besides the improvements and the reduction in the number of features to represent the facial expressions, contrary to other state-of-the art approaches, the proposed approach does not rely on fiducial points for feature extraction. This makes the feature extraction more reliable because the correct feature extraction does not depend on the correct location of the fiducial points which are either manually assigned or using automatic procedures that are error-prone. Furthermore, the process is much simpler and does not require human intervention.

The statistical tests have shown that the proposed approach achieves a superior or equivalent performance relative to other state-of-the-art approaches while relieving the computational effort. The experimental results have shown that the proposed approach compares favorably against other approaches in terms of both accuracy and in the number of processed features.

A main drawback of the proposed method is that it requires several classifiers. Therefore we have compared the total number of features evaluated during the classification as a measure of computational effort, in particular relative to the approaches that use SVMs. However, most of the methods employ multi-class SVMs [10], [11], [20] and that implies in at least seven binary SVMs if the one-against-all strategy is adopted.

Acknowledgment

This research has been supported by The National Council for Scientific and Technological Development (CNPq) grant 311.832/2013-0.

References

[1] M. Mansano, L. E. S. Oliveira, A. L. Koerich, and A. S. Britto Jr., "2d principal component analysis for face and facial-expression recognition," *Computing in Science & Engineering*, vol. 13, no. 3, pp. 9–13, may 2011.

[2] A. Dhall, "Context Based Facial Expression Analysis in the Wild," in *3rd ACM International Conference on Multimedia Retrieval*, sep 2013, pp. 636–641.

[3] M. Hussain, S. A. Khan, N. Ullah, N. Riaz, and M. Nazir, "Computationally Efficient Invariant Facial Expression Recognition," *Res. J. Recent Sci.*, vol. 3, no. 2, pp. 61–68, 2014.

[4] A. Mehrabian, "Communication Without Words," *Psychol. Today*, vol. 2, no. 4, pp. 53–56, 1968.

[5] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, feb 1971.

[6] X. Wang, C. Jin, W. Liu, M. Hu, L. Xu, and F. Ren, "Feature Fusion of HOG and WLD for Facial Expression Recognition," *Syst. Integr. (SII), 2013 IEEE/SICE Int. Symp.*, pp. 227–232, 2013.

TABLE III

PERFORMANCES OF THE GLOBAL AND THE PAIRWISE APPROACHES ON THE THREE DATASETS USING SVM CLASSIFIERS AND DIFFERENT DIMENSIONALITY REDUCTION APPROACHES.

Feature Set	Approach	Correct classification rate (%)								
		JAFFE			CK			TFEID		
		global	global FS	pairwise FS	global	global FS	pairwise FS	global	global FS	pairwise FS
WLD	Full	90.05	—	—	89.40	—	—	94.78	—	—
WLD	filter CFS	—	81.99	98.10	—	91.08	96.14	—	97.39	98.88
WLD	hybrid IG+SVM	—	81.52	97.63	—	91.08	96.14	—	96.64	98.51
WLD	hybrid KW+SVM	—	89.10	**99.05**	—	90.60	96.87	—	97.01	98.88
LBP	Full	91.94	—	—	91.81	—	—	93.28	—	—
LBP	filter CFS	—	94.31	98.10	—	92.05	97.83	—	95.90	99.25
LBP	hybrid IG+SVM	—	92.89	98.10	—	92.05	**98.07**	—	95.15	99.25
LBP	hybrid KW+SVM	—	95.26	**99.05**	—	90.60	96.87	—	96.27	**99.63**

TABLE IV

COMPARISON OF THE PROPOSED APPROACH TO THE PREVIOUS WORKS REPORTED ON JAFFE AND CK DATASETS

Reference	Dataset	# of features	Classification rate (%)
Bashar et al. [8]	CK	21,504	92.2
Zavaschi et al. [11]	CK	2,478 + 160	99.4
Zavaschi et al. [11]	JAFFE	2,478 + 160 + 300	96.2
Wang et al. [6]	JAFFE	1,440 + 2,560	93.9
Liu et al. [9]	JAFFE	864	96.1
Zhang et al. [16]	JAFFE	18	71.0
Zhang et al. [16]	CK	18	91.0
Proposed Approach	CK	2,268	98.1
Proposed Approach	JAFFE	1,008	99.1

[7] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *Computer Vision and Pattern Recognition Conference*, vol. 1, 2001, pp. 511–518.

[8] F. Bashar, A. Khan, F. Ahmed, and M. H. Kabir, "Robust facial expression recognition based on median ternary pattern (MTP)," *International Conference on Electrical Information and Communication Technology*, pp. 1–5, feb 2013.

[9] S. Liu, Y. Zhang, and K. Liu, "Facial expression recognition under partial occlusion based on Weber Local Descriptor histogram and decision fusion," in *33rd Chinese Control Conference*, Hanjing, China, Jul 2014, pp. 4664–4668.

[10] R. Verma and M. Y. Dabbagh, "Fast Facial Expression Recognition Based On Local Binary Patterns," in *26th Annual IEEE Canadian Conference on Electrical and Computer Engineering*, Regina, Canada, 2013, pp. 1–4.

[11] T. H. H. Zavaschi, A. S. Britto Jr., L. E. S. Oliveira, and A. L. Koerich, "Fusion of feature sets and classifiers for facial expression recognition," *Expert Systems with Applications*, vol. 40, no. 2, pp. 646–655, feb 2013.

[12] J. K. Pontes, A. S. Britto Jr., C. Fookes, and A. L. Koerich, "A flexible hierarchical approach for facial age estimation based on multiple features," *Pattern Recognition*, vol. 54, pp. 34–51, Jun 2016.

[13] Z. Huang and F. Ren, "Facial Expression Recognition basead on Active Appearance Model And Scale-Invariant Feature Transform," in *2013 IEEE/SICE International Symposium on System Integration*, Kobe, Japan, 2013, pp. 94–99.

[14] C. Martin, U. Werner, and H.-M. Gross, "A real-time facial expression recognition system based on Active Appearance Models using gray images and edge images," *8th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1–6, 2008.

[15] M. Kyperountas, A. Tefas, and I. Pitas, "Pairwise facial expression classification," *IEEE International Workshop on Multimedia Signal Processing*, pp. 1–4, oct 2009.

[16] Z. Zhang, C. Fang, and X. Ding, "Facial expression analysis across databases," *International Conference on Multimedia Technology*, pp. 317–320, jul 2011.

[17] S. Meher and P. Maben, "Face recognition and facial expression identification using PCA," in *Adv. Comput. Conf. (IACC), 2014 IEEE Int.*, 2014, pp. 1093–1098.

[18] X. Li, Q. Ruan, G. An, and Y. Jin, "Automatic 3D facial expression recognition based on polytypic Local Binary Pattern," in *2014 12th Int. Conf. Signal Process.*, 2014, pp. 1030–1035.

[19] S. Dinakaran and P. R. J. Thangaiah, "Comparative Analysis of Filter-Wrapper Approach for Random Forest Performance on Multivariate Data," *2014 Int. Conf. Intell. Comput. Appl.*, pp. 174–178, 2014.

[20] T. H. H. Zavaschi, L. E. S. Oliveira, and A. L. Koerich, "Facial Expression Recognition Using Ensemble of Classifiers," in *36th International Conference on Acoustics, Speech and Signal Processing (ICASSP2011)*, Prague, Czech Republic, 2011, pp. 1489–1492.

[21] Y. Liu, R. Wang, and Y. S. Zeng, "An improvement of one-against-one method for multi-class support vector machine," *Proc. Sixth Int. Conf. Mach. Learn. Cybern. ICMLC 2007*, vol. 5, no. August, pp. 2915–2920, 2007.

[22] M. Huang, Z. Wang, and Z. Ying, "Facial expression recognition using Stochastic Neighbor Embedding and SVMs," *Proc. 2011 Int. Conf. Syst. Sci. Eng.*, no. June, pp. 671–674, 2011.

[23] H. Ghaderi and P. Kabiri, "Fourier transform and correlation-based feature selection for fault detection of automobile engines," *AISP 2012 - 16th CSI Int. Symp. Artif. Intell. Signal Process.*, no. Aisp, pp. 514–519, 2012.

[24] L. Patil and M. Atique, "A novel feature selection based on information gain using WordNet," in *Sci. Inf. Conf.*, London, 2013, pp. 625–629.

[25] M. Last, H. Bunke, and a. Kandel, "A feature-based serial approach to classifier combination," *Pattern Anal. Appl.*, vol. 5, pp. 395–398, 2002.