

Facial Expression Recognition: A Cross-Database Evaluation of Features and Classifiers

Francisco Edvan Chaves, Thelmo Pontes de Araujo, José Everardo Bessa Maia
State University of Ceará – UECE, Brazil
edvanchaves@gmail.com, thelmo.araujo@uece.br, jose.maia@uece.br



ABSTRACT: Many approaches in both feature extraction and classification have been proposed in order to build a robust automatic facial expression recognition system. The chosen features and classifiers are usually compared in limited scenarios. In this paper, 5 feature extractors (3 variations of Gabor filters, Local Binary Patterns, and Discrete Cosine Transform) and 4 classifiers (K-Nearest Neighbors, Support Vector Machine, Radial Basis Function Neural Network, and Naive Bayes) were combined and applied to three different datasets: JAFFE, Yale, and CK+. The combinations of feature extractors and classifiers were compared in more robust settings, being evaluated in each dataset separately and in 3 cross-database settings, to verify the technique's generalization power. All experiments were performed in two validation scenarios: in the first one, the system tries to recognize the facial expression of a person already known to it in the training phase, using an image not present in the training set (leave-one-out); in the other scenario, all images of a certain individual are in the testing set, so the system tries to recognize the facial expression of a person unknown to it in the training phase (leave-one-subject-out). Accuracy results, as well as computational times, are presented, suggesting the combination of feature extractors and classifiers more suited for generalization.

Keywords: Facial Expressions, Feature Extraction, Classification, Cross-Database Evaluation

Received: 12 August 2018, Revised 15 October 2018, Accepted 5 November 2018

DOI: 10.6025/jic/2019/10/1/34-45

© 2019 DLINE. All Rights Reserved

1. Introduction

Facial expression refers to the appearance of the human face as a way to transmit emotions. Ekman and Friesen [4] propose six universal facial expressions: happiness, anger, sadness, surprise, disgust, and fear. An automatic Facial Expression Recognition (FER) system uses a computer to read a digital image of a human face and to attribute one of those six facial expressions to it. Video streams of faces may also be used to enrich the discriminative power of a FER system.

An active research in FER has taken place nowadays and many applications are envisioned in areas such as Human- Computer Interaction, Ubiquitous Computing, Internet of Things, lie detection, games, and so on. In [7], for example, the recognition of basic emotions from facial expressions in young people with autism spectrum disorder is studied. The assumption behind automatic FER is the existence of certain regularities in patterns of facial muscle configurations and textures and the relative positions of the

human face parts whenever the person has an emotional reaction. The continuous variability of these features and the diversity of human faces make the building of such a system a challenging task. Some commercial softwares may perform FER, but their results are not robust enough, leaving much room for improvement.

FER may be approached as a statistical pattern recognition problem, having four stages: preprocessing, feature extraction, dimensionality reduction, and classification. Although a great performance difference do exist among classifiers (e.g., linear and non-linear classifiers), it is well established that the discriminative power of the features is the main determinant of the classifier performance [26].

Deng and colleagues [3] use a Gabor filter bank with 40 filters (5 scales and 8 orientations) for feature extraction. Images of each person in the JAFFE [10] database were used for both training and testing phases. The best accuracy (97.33%) was obtained using 12 Gabor filters, PCA and LDA for dimensionality reduction, and Euclidean distance for classification.

In 2012, Gu and colleagues [8] also applied a Gabor filter bank (with 24 filters: 3 scales and 8 orientations) to JAFFE and CK [13] databases. The filtered blocks were then mapped to a grid that simulates the human visual cortex. The mean accuracy for JAFFE was 89:67%, and 91:51% for the CK database. The authors also used a cross-methodology, training the system with the JAFFE images and testing it with CK images: They obtained a mean accuracy of 54:05%. The mean accuracy was a little higher (55:87%) when the system was trained/tested the other way around.

Mayer et al. [15] performed a cross-database evaluation of the Candide-III face model in three datasets: CK+, MMI, and FEEDTUM. Their best-scenario experiment achieved 66:2% of accuracy.

Owusu and colleagues [20] used the Viola-Jones algorithm [25] to detect faces on images from JAFFE and Yale [28] databases. 40 Gabor filters were applied in the feature extraction phase and a Multilayer Perceptron (MLP) was used in the classification phase. The authors obtained a high accuracy rate (96:83%), but images of all persons were in the training set (i.e., all persons were known by the system).

Local Binary Patterns (LBP) features are used in [22, 1, 16]. In [22], the authors selected the most relevant features from the histograms of uniform Local Binary Patterns (LBP) using a method called Boosted-LBP. A mean accuracy of 92:6% were obtained in the CK database. Abdulrahman and colleagues [1] also used LBP (combined with Gabor filters) and achieved 90:0% of accuracy in the JAFFE database.

Discrete Cosine Transform (DCT) is another common feature extractor. In [14], for example, a subset of DCT coefficients obtained from the application of DCT to the difference between the images with non-neutral expressions and the neutral ones was used to achieve a mean accuracy of 93:75% on a their own database, which consisted of 60 persons (40 for training and 20 for testing).

Bhadu and colleagues [2] also extracted DCT coefficients from JAFFE images, using the Adaboost classifier to achieve an accuracy of 71%. Gabor filters and Daubechies wavelets did not improve the mean accuracy achieved by the Adaboost classifier. Wavelet domain is also explored by [17], using histogram of oriented gradients; the proposed system was evaluated in CK+, JAFFE and Yale facial expression datasets and outperformed some existing methods.

Ghimire et al. [5] used an elastic bunch graph matching (EBGM) algorithm the Kanade-Lucas-Tomaci (KLT) tracker to select the most relevant features to be extracted by a multi-class AdaBoost and an extreme learning machine (ELM) classification scheme. The FER system was evaluated in CK+, MMI and MUG facial expression data sets.

In [29], [12], and [24], the authors used Convolutional Neural Networks (CNN) to perform the facial recognition task. In this paper, a cross-database evaluation was performed with the combination of 5 feature extractors (3 variations of Gabor filters, Local Binary Patterns, and Discrete Cosine Transform) and 4 classifiers (K-Nearest Neighbors, Support Vector Machine, Radial Basis Function Neural Network, and Naive Bayes). All 20 combined techniques were applied to 3 datasets: JAFFE, Yale, and CK+, and results were presented for both intra-database (training and testing with the same dataset) and inter-database (training with one dataset and testing with the other two) experiments, in order to verify the technique's generalization strength. In each experiment, two validation techniques were considered: in the leave-one-out validation, the system tries to classify the facial expression of a person known to it in the training phase; In the other validation technique, the system tries to recognize the facial

expression of a person unknown to it in the training phase.

This work continues with Section 2 presenting the techniques for feature extraction, dimensionality reduction, and classification with their respective methodology, as well as the datasets. Next, in Section 3, experiments and results are shown and discussed. Section 4 concludes the paper.

2. Methodology

There are many local and holistic approaches for automatic FER systems. Local approaches, such as LBP, Gabor filters, and DCT, have shown better performance than holistic ones, leading us to choose them in this work.

2.1 Feature Extraction

Three features were extracted from the images in the databases: Gabor filters, Local Binary Patterns (LBP), and the Discrete Cosine Transform (DCT).

2.1.1 Gabor Filters

A Gabor filter is basically a Gaussian function modulated by a complex exponential and is known to model the mammal visual cortex [23]. Here we used the same formulation as in [3]:

$$g(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x')^2 + (y')^2}{2\sigma^2}\right) \left[\exp(i\omega x') - \exp\left(-\frac{\omega^2\sigma^2}{2}\right) \right],$$

where $z = (x, y)$ is a point in \mathbb{R}^2 , $x' = x \cos \theta + y \sin \theta$, $y' = -x \sin \theta + y \cos \theta$, θ is the orientation, ω is the frequency, σ^2 is the variance of the Gaussian kernel, and i is the imaginary unit.

We followed a common practice [3, 20] and used 40 Gabor filters (5 scales and 8 orientations). The frequency is taken to be $\omega = \omega_{max} \lambda^{-(m-1)}$, with $\omega_{max} = \pi/2$, $\lambda = \sqrt{2}$, and $m = 1, 2, \dots, 5$. The filter orientations are given by $\theta_n = (n - 1)/8$, with $n = 1, 2, \dots, 8$ (see Figure 1).

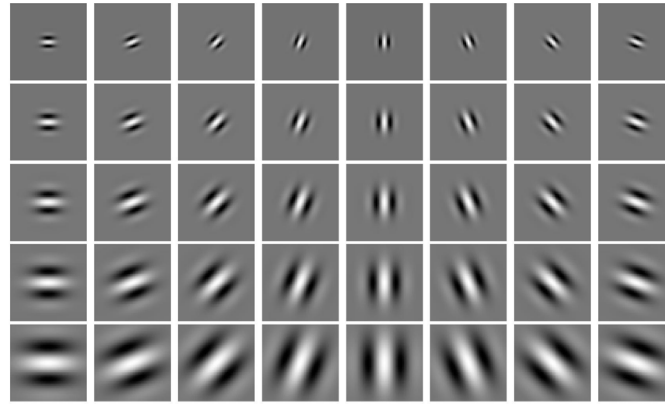


Figure 1. Gabor filter bank: 5 scales and 8 orientations (real part)

After the convolution of each Gabor filter, the filtered (magnitude) images are concatenated into a feature vector, which has dimension 491 520 (= 40 96 128). Due to this large dimension, two techniques for dimensionality reduction were applied to the resulting feature vector (PCA and LDA), which are explained on Section 2.2.

2.1.2 Local Binary Pattern

A Local Binary Pattern (LBP) of each pixel in a gray scale image may be obtained by assigning 0 or 1 to each of its 8 neighbors: 0 if the neighbor's gray level is lower than the central pixel's level; and 1 otherwise. An 8-bit binary number is formed by concatenating the clockwise sequence of bits, and then converted to a decimal number (between 0 and 255), which now represents the central pixel in the transformed image, called LBP image (see Figure 2).

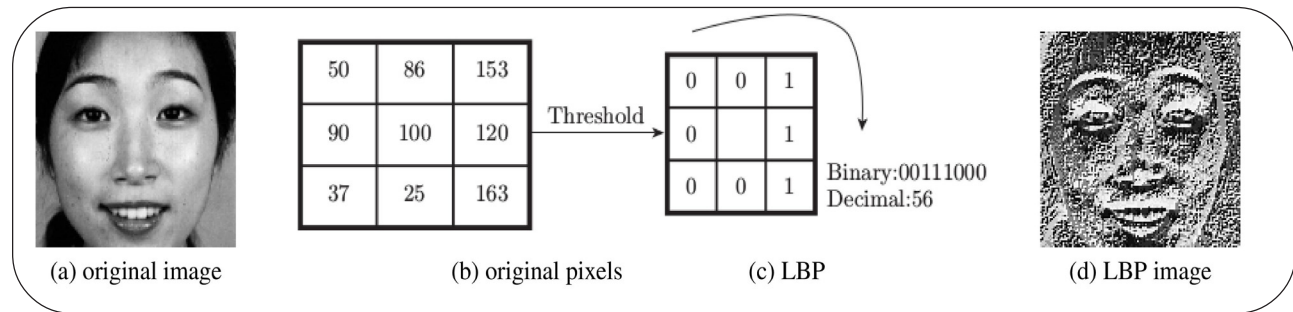


Figure 2. Example of LBP

A local binary pattern is considered uniform if the corresponding binary number has at most two bit changes; if it has more than two bit changes, it is called non-uniform. For example, 00011100 is a uniform pattern, but 011100011 is a non-uniform pattern.

Once the LBP image is obtained, its histogram is used as the feature vector of the original image. As histograms lack localization, the original image might be partitioned in blocks in order to preserve some localization: each segment of the feature vector corresponds to a certain region of the original image. This plays an important role in a FER system [22]. Here images were partitioned into 48 blocks of size 16 16 and histograms have 58 bins, resulting in a feature vector of dimension 2784.

2.1.3 Discrete Cosine Transform

The Discrete Cosine Transform (DCT) is frequently used in signal processing and data compression mainly for its energy compaction property. DCT-II [6] is used in this paper:

$$F(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} f(x, y) \cos\left(\frac{(2x+1)u\pi}{2w}\right) \cos\left(\frac{(2y+1)v\pi}{2h}\right),$$

where w and h are the image dimensions, and

$$\alpha(u) = \begin{cases} \sqrt{1/w}, & \text{if } u = 0, \\ \sqrt{2/w}, & \text{otherwise.} \end{cases}$$

DCT is usually performed on small image blocks (48 blocks of size 16 x 16, in our case), and since the most important coefficients are in the left upper corner of the transformed image, the feature vector is formed by concatenating only the 16 (= 4 x 4) DCT coefficients. The resulting feature vector has dimension 768.

2.2 Dimensionality Reduction

Two techniques were used to reduce the dimensions of Gabor feature vectors: Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

PCA aims to find an orthogonal (non-correlated) basis that spans the same linear subspace as the one spanned by the feature vectors. This orthogonal basis is ordered in increasing variability, so keeping only the first vectors (the number depends on each situation) one may accounts for most of the data variability [11].

LDA, unlike PCA, considers the within-class and between-class variabilities finding the linear combination of features vectors that separates classes as much as possible [26]. Fisher's criterion aims to maximize the between-class to within-class ratio.

In this paper, we considered 3 sets of Gabor feature vectors: In the first set a simple sampling technique (with a sampling factor of 4) was applied to the feature vector to reduce its dimension to 30720 (40 x 24 x 32); In the second one, PCA was applied to further reduce the feature vector dimension, using only the first 80 principal components; In the third set, LDA was applied to the Gabor+PCA feature vectors. We refer to these sets as Gabor, Gabor/PCA, and Gabor/LDA, respectively.

2.3 Classifiers

Once the feature images are obtained, one must proceed with classification. Four kinds of classifiers were used in the work, both linear and non-linear ones, mainly for their robustness and popularity in the literature.

K-Nearest Neighbor (K-NN) is a supervised, non-parametric classification technique that compares the test sample with its K nearest neighbors, and assigns the predominant class of the K neighbors to it [18]. The Euclidean distance was the metric used in this work, with $K = 1$.

Support Vector Machine (SVM) is a non-parametric classifier which aims to find a decision boundary with maximum distance between the classes. SVM maps the data into a higher dimension in which the classes are linearly separable [18, 9]. A Sequential Minimal Optimization (SMO) algorithm, proposed by Platt and colleagues [21], is used here to train SVM classifiers.

Radial Basis Function Neural Networks (RBFNN) use a radial function, such as the Gaussian function used here, as the activation function in the hidden layer [9]. RBFNN has a non linear feed-forward architecture to perform supervised training.

In order to contrast with the non-linear classifiers used here, we chose Naive Bayes as a linear classifier. It uses Bayes Theorem to classify patterns by making the strong assumption of feature independence. Even if the features are clearly dependent, Naive Bayes shows good accuracy in classification problems [26, 18].

2.4 Datasets

Three datasets were used: JAFFE, Yale, and CK+. JAFFE (Japanese Female Facial Expression) dataset [10] has 213 face images of 10 Japanese female individuals with at least 2 of each expression: happiness (HA), sadness (SA), surprise (SU), disgust (DI), anger (AN), fear (FE), and neutral (NE) (see Figure 3). The original images are 256 x 256 pixels and all of them were used in the experiments.



Figure 3. Examples of facial expressions in JAFFE dataset

60 images of 15 individuals, with only one image for each of 4 different facial expressions were selected from Yale dataset [28]. Figure 4 shows examples for each facial expression: happiness (HA), sadness (SA), surprise (SU), and neutral (NE).

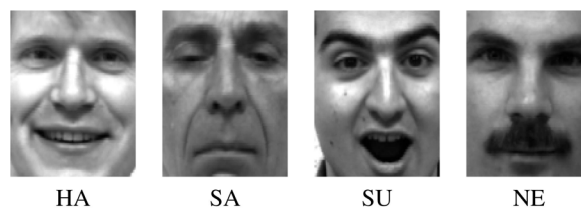


Figure 4. Examples of facial expressions in Yale dataset

CK+ (extended Cohn-Kanade) dataset [13] has 593 images of 123 different persons. The sequence of images varies from person to person (from 10 to 60), starting with a neutral expression and ending with the maximum corresponding expression. In our experiments, the 327 images labeled with facial expressions happiness (HA), sadness (SA), surprise (SU), disgust (DI), anger (AN), fear (FE), and contempt (CO) were used (see Figure 5).

3. Experiments, Results, and Discussion

The experiments were realized in a desktop with Intel Core i5 chipset, 6 MB of RAM, and AMD Radeon HD 512 MB graphics card.

Preprocessing, feature extraction, and dimensionality reduction were implemented in C++ language using OpenCV computer vision library [19]. Tests were performed in JAVA with WEKA data mining library [27].

All images from the three datasets were preprocessed using the Viola-Jones method to locate the faces in the images, which were then clipped and resized to 128 96 pixels. Color images were converted to 8-bit gray scale.

Gabor, LBP, and DCT features were then extracted, and, in each set of feature vectors (including the variations Gabor/PCA and Gabor/LDA), four classifiers were applied: K-Nearest Neighbor (K-NN), Support Vector Machine (SVM-SMO), Radial Basis Function Neural Network (RBFNN), and Naive Bayes.

3.1 Experiments with JAFFE database

Two validation techniques were used with the JAFFE database: leave-one-out and leave-one-subject-out. In the first one, in each of the 213 runs, one image was selected for testing and the 212 others for training. In this case, the system tries to recognize the facial expression of a person known to the system in the training phase. In the leave-one-subject-out technique, the whole set of images of a certain individual was selected for testing in each of the 10 runs. Therefore, the system tries to recognize the facial expression of an unknown person.

Table 1 shows the accuracy results for both validation experiments with JAFFE dataset. The SVM-SMO classifier using Gabor features attained the best performance consistently: 96:71% for known subjects and 69:48% for unknown subjects.

	Leave-one-out				Leave-one-subject-out			
	K-NN	SVM	RBF	Bayes	K-NN	SVM	RBF	Bayes
Gabor	92.49	96.71	80.75	81.22	49.77	69.48	59.15	61.97
Gabor/PCA	90.14	93.43	86.38	79.81	49.30	69.01	53.05	53.52
Gabor/LDA	93.90	94.84	96.24	94.84	66.67	66.67	68.08	67.61
LBP	87.32	86.38	62.91	67.14	49.30	63.38	50.70	52.58
DCT	88.73	91.55	73.24	69.48	39.44	55.87	50.23	52.58

Table 1. Accuracy (%) obtained for the JAFFE dataset

The confusion matrices for the winning combination in both cases are shown on Tables 2 and 3. For the leave-oneout experiment, Gabor+SVM performed well, with worst accuracy of 93:33%. As for the leave-one-subject-out, the technique obtained a poor accuracy rate only when classifying fear expression (37:50%).

	HA	SA	SU	DI	AN	FE	NE
HA	96.77	3.23	0.00	0.00	0.00	0.00	0.00
SA	3.23	93.55	0.00	0.00	0.00	3.23	0.00
SU	3.33	0.00	93.33	0.00	0.00	3.33	0.00
DI	0.00	0.00	0.00	96.55	3.45	0.00	0.00
AN	0.00	0.00	0.00	3.33	96.6	70.00	0.00
FE	0.00	0.00	0.00	0.00	0.00	100.00	0.00
NE	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Table 2. Confusion matrix (%) for JAFFE dataset using Gabor + SVM-SMO (leave-one-out)

	HA	SA	SU	DI	AN	FE	NE
HA	70.97	6.45	12.90	0.00	0.00	0.00	9.68
SA	3.23	61.29	3.23	16.13	3.23	9.68	3.23
SU	3.33	0.00	83.33	0.00	0.00	6.67	6.67
DI	0.00	6.90	0.00	75.86	17.24	0.00	0.00
AN	3.33	10.00	0.00	3.33	80.00	0.00	3.33
FE	0.00	18.75	15.62	18.75	0.00	37.50	9.38
NE	0.00	0.00	3.33	0.00	16.67	0.00	80.00

Table 3. Confusion matrix (%) for JAFFE dataset using Gabor + SVM-SMO (leave-one-subject-out)

3.2 Experiments with Yale Database

Experiments with the Yale dataset were performed with two validation techniques: leave-one-out and leave-one-groupout. In the leave-one-group-out technique, the following procedure was repeated 5 times: all images of 3 different persons were used for testing, while the remaining images were used for training. The 5 testing sets have no common individuals.

Table 4 shows the accuracy results for both validation experiments with Yale dataset. RBFNN classifier using Gabor/LDA features attained the best performance (70:00%) for the leave-one-out experiment; as for the leave-onegroup- out experiment, the SVM-SMO classifier using LBP features was far better (78:33%) than the other pairs.

	Leave-one-out				Leave-one-group-out			
	-NN	SVM	RBF	Bayes	K-NN	SVM	RBF	Bayes
Gabor	1.67	46.67	30.00	31.67	61.67	63.33	63.33	61.67
Gabor/PCA	1.67	40.00	30.00	21.67	48.33	66.67	60.00	61.67
Gabor/LDA	65.00	58.33	70.00	68.33	60.00	66.66	66.67	61.67
LBP	1.67	51.67	38.33	25.00	58.33	78.33	45.00	63.33
DCT	10.00	40.00	21.67	33.33	46.67	58.33	48.33	51.67

Table 4. Accuracy (%) obtained for the Yale dataset

The confusion matrices for the winning combination in both cases are shown on Tables 5 and 6. The best accuracies are obtained when classifying happiness (HA) and surprise (SU).

	HA	SA	SU	NE
HA	100.00	0.00	0.00	0.00
SA	6.67	53.33	0.00	40.00
SU	0.00	20.00	73.33	6.67
NE	0.00	40.00	6.67	53.33

Table 5. Confusion matrix (%) for Yale dataset using Gabor/LDA + RBF (leave-one-out)

	HA	SA	SU	NE
HA	93.33	0.00	0.00	6.67
SA	6.67	60.00	0.00	33.33
SU	0.00	0.00	100.00	0.00
NE	0.00	40.00	0.00	60.00

Table 6. Confusion matrix (%) for Yale dataset using LBP + SVM-SMO (leave-one-group-out)

3.3 Experiments with CK+ Database

Again, experiments with the CK+ dataset were performed with leave-one-out and leave-one-group-out validation techniques. Ten testing sets were formed with all images of 12 or 13 different persons, with no common individuals in the testing sets of each validation run.

Table 7 shows the accuracy results for both validation experiments with CK+ dataset. The SVM-SMO classifier using Gabor and Gabor/LDA features attained again the best performance consistently: 92:05% for known subjects and 93:27% for unknown subjects.

	Leave-one-out				Leave-one-group-out			
	K-NN	SVM	RBF	Bayes	K-NN	SVM	RBF	Bayes
Gabor	48.62	91.44	80.43	74.92	83.49	93.27	82.57	80.12
Gabor/PCA	37.72	89.60	80.12	83.18	76.15	91.74	84.10	88.07
Gabor/LDA	89.60	92.05	90.83	90.83	92.05	91.74	91.74	92.97
LBP	38.23	88.69	74.62	75.23	78.29	90.83	75.84	81.65
DCT	53.52	82.87	76.76	74.92	74.92	85.32	78.29	78.90

Table 7. Accuracy (%) obtained for the CK+ dataset

Tables 8 and 9 show the confusion matrices for the leave-one-out and leave-one-group-out experiments, respectively.

Happiness (HA), surprise (SU), disgust (DI), and anger (AN) expressions attained the best accuracies.

	AN	CO	DI	FE	HA	SA	SU
AN	93.33	0.00	2.22	0.00	0.00	4.44	0.00
CO	22.22	72.22	0.00	0.00	0.00	5.56	0.00
DI	0.00	0.00	98.3	1 0.00	1.69	0.00	0.00
FE	8.00	0.00	0.00	84.00	4.00	0.00	4.00
HA	0.00	1.45	0.00	0.00	98.55	0.00	0.00
SA	35.71	0.00	0.00	0.00	0.00	60.71	3.57
SU	0.00	1.20	0.00	0.00	0.00	0.00	98.80

Table 8. Confusion matrix (%) for CK+ dataset using Gabor/LDA + SVM-SMO (leave-one-out)

	AN	CO	DI	FE	HA	SA	SU
AN	95.56	2.22	2.22	0.00	0.00	0.00	0.00
CO	11.11	72.22	0.00	0.00	0.00	5.56	11.11
DI	1.69	0.00	96.61	0.00	1.69	0.00	0.00
FE	8.00	0.00	0.00	84.00	0.00	0.00	8.00
HA	0.00	0.00	0.00	0.00	100.00	0.00	0.00
SA	17.86	0.00	0.00	0.00	0.00	71.43	10.71
SU	0.00	1.20	0.00	0.00	0.00	0.00	98.80

Table 9. Confusion matrix (%) for CK+ dataset using Gabor + SVM-SMO (leave-one-group-out)

3.4 Cross-database Experiments

In order to evaluate generalization robustness of the feature/classifier pairs tested in this paper, cross-database experiments were realized, i.e., one database was used for training and the other two for testing.

Since the datasets used here do not have the same set of facial expressions, the cross-database experiments were performed with only four facial expressions: happiness (HA), sadness (SA), surprise (SU), and neutral (NE). 33 images from CK+ database with neutral facial expression were labeled and included in the CK+ subset for the cross database experiments.

Tables 10, 11, and 12 show, respectively, the accuracy results for the three cross-database experiments: training with JAFFE database and testing with Yale and CK+ datasets; training with Yale database and testing with JAFFE and CK+ datasets; training with CK+ database and testing with JAFFE and Yale datasets.

	Yale				CK+			
	K-NN	SVM	RBF	NB	K-NN	SVM	RBF	NB
Gabor	46.67	41.67	43.33	38.33	64.32	74.18	72.77	52.58
Gabor/PCA	41.67	41.67	38.33	35.00	48.36	66.67	53.99	66.67
Gabor/LDA	40.00	28.33	41.67	43.33	46.48	44.13	39.91	47.89
LBP	40.00	45.00	46.67	45.00	59.62	76.06	54.46	66.20
DCT	33.33	38.33	25.00	28.33	43.19	42.72	48.83	47.42

Table 10. Cross-database experiment: training with JAFFE dataset

Results on Tables 10, 11, and 12 confirm the consistent superiority of the variants of Gabor features over DCT features. LBP features also showed good performance often outperforming DCT features and sometimes showed better results than Gabor features. As for the classifiers, SVM-SMO was consistently better than the other classifiers, but was surpassed by RBFNN classifier sometimes. The combination of Gabor features and SVM-SMO classifier proved to be robust under generalization, although even its best performance (96:71%) may be surpassed by a fine tuned state-of-the-art classifier applied to a single database.

Table 13 shows the confusion matrix for the best result on the cross-database experiments: Training with Yale database and testing with CK+ dataset.

	JAFPE				CK+			
	K-NN	SVM	RBF	NB	K-NN	SVM	RBF	NB
Gabor	55.74	36.89	56.56	34.43	80.75	88.73	78.87	76.53
Gabor/PCA	53.28	34.43	24.59	25.41	75.59	82.16	66.20	74.65
Gabor/LDA	33.61	37.70	25.41	44.26	77.93	80.75	61.03	77.00
LBP	25.41	24.59	25.41	26.23	65.26	80.75	53.52	60.09
DCT	49.18	24.59	37.70	39.34	55.87	80.28	76.06	79.81

Table 11. Cross-database experiment: training with Yale dataset

	JAFPE				Yale			
	K-NN	SVM	RBF	NB	K-NN	SVM	RBF	NB
Gabor	40.16	45.08	43.44	39.34	55.00	73.33	56.67	56.67
Gabor/PCA	45.08	40.16	43.44	40.98	46.67	63.33	38.33	48.33
Gabor/LDA	41.80	42.62	43.44	44.26	65.00	61.67	58.33	68.33
LBP	39.34	45.90	40.16	28.69	56.67	66.67	68.33	63.33
DCT	28.69	31.15	31.15	35.25	41.67	65.00	56.67	56.67

Table 12. Cross-database experiment: training with CK+ dataset

	HA	SA	SU	NE
HA	100.00	00.00	00.00	00.00
SA	00.00	35.71	21.43	42.86
SU	00.00	00.00	98.80	01.20
NE	00.00	00.00	15.15	84.85

Table 13. Confusion matrix (%) for best accuracy of cross-database experiments: training with Yale dataset, testing with CK+ dataset

	Extraction	K-NN	SVM	RBF	NB
Gabor	120.37	192.9	07.68	19.61	89.57
Gabor/PCA	145.06	0.58	0.08	0.11	0.50
Gabor/LDA	148.02	0.19	0.07	0.07	0.11
LBP	32.00	9.49	0.69	1.95	7.18
DCT	4.00	3.60	0.70	0.50	2.50

Table 14. Processing time (in ms)

Another relevant issue when looking for a good classification technique is the computational time spending on the two main phases of the techniques shown here: feature extraction and classification. As Table 14 shows, although Gabor features generalize well, the time to extract them is much higher than the time to extract the other features. Without PCA or LDA to reduce the dimensions of the Gabor feature vectors, time for classification is also high. DCT proved to be very useful if one needs real time feature extraction, specially if used together with SVM-SMO or RBFNN classifiers.

The experiments were planned and performed to compass two dimensions of comparison: Within database versus cross-database and leave-one-out versus leave-one-subject-out. In the first case, six within database and six crossdatabase experiments were performed. In the second case, six leave-one-out and six leave-one-subject-out experiments were also performed.

Results show the SVM classifier associated with Gabor features surpassing the other classifier-feature pairs in 9 of the 12 experiments, which makes it the most recommended scheme when accuracy is crucial. On the other hand, if a compromise between accuracy and processing time is important, one may use the results on Table 14 to choose the most suitable pair.

4. Conclusion

Although Gabor, LBP, and DCT features are presented in the literature as most promising features, they are usually compared in partial, limited scenarios. This work aimed to compare these features in more robust scenarios: We evaluated 20 combinations of 5 feature extractors and 4 classifiers in 3 different datasets, showing results for both intra-database and inter-database experiments.

Even when dealing with the same database, variations in the feature extraction may cause great differences in the accuracy results, as can be seen in [3] and [8], for example. Therefore, it was expected no consistent superiority from a particular technique in the experiments.

Nevertheless, the combination of Gabor features and SVM classifier proved to be a robust technique for facial expression classification in both intra-database and inter-database scenarios. If computational time is of relevance, the combination of LBP features and SVM classifier is better suited to the task.

The results shown here corroborate the precariousness of image level features on capturing and generalizing facial expressions. This weakness points to the need of searching for more holistic features (which are, probably, the way used by the human recognition system), which is the direction of our future work.

References

- [1] Abdulrahman, Muzammil., Tajuddeen, R., Gwadabe, Abdu, Fahad J., Eleyan, Alaa (2014). Gabor wavelet transform based facial expression recognition using PCA and LBP. *In: 2014 22nd Signal Processing and Communications Applications Conference (SIU)*, Trabzon, Turkey, April 23-25, 2014, p. 2265–2268.
- [2] Bhadu, Aruna., Tokas, Rajbala., Vijay Kumar. (2012). Facial expression recognition using DCT, Gabor and wavelet feature extraction techniques. *International Journal of Engineering and Innovative Technology*, 2 (1).
- [3] Deng, Hong-Bo., Jin, Lian-Wen., Zhen, Li-Xin., Huang, Jian-Cheng. (2005). A new facial expression recognition method based on local Gabor filter bank and PCA plus LDA. *International Journal of Information Technology*, 11(11) 86–96.
- [4] Ekman, Paul., Friesen, Wallace V.(1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17 (2) 124–129.
- [5] Ghimire, Deepak., Lee, Joonwhoan., Li, Ze-Nian., Jeong, Sunghwan., (2017). Recognition of facial expressions based on salient geometric features and support vector machines. *Multimedia Tools and Applications*, 76 (6) 7921–7946.
- [6] Rafael, C., Gonzalez., Richard, E. (2007). *Woods. Digital Image Processing*. Prentice Hall, 3 edition, August.
- [7] Griffiths, Sarah., Jarrold, Christopher ., Penton-Voak, Ian S., Woods, Andy T., Skinner., Andy L., Munafò, Marcus R. (2017). Impaired recognition of basic emotions from facial expressions in young people with autism spectrum disorder: assessing the importance of expression intensity. *Journal of Autism and developmental disorders*, p. 1–11.

- [8] Gu, Wenfei., Xiang, Cheng., Venkatesh, YV., Huang, Dong., Lin, Hai. (2012). Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Pattern Recognition*, 45(1), 80–91.
- [9] Haykin, Simon S. (2009). *Neural Networks and Learning Machines*, 3. Pearson Education Upper Saddle River.
- [10] JAFFE. (1998). The japanese female facial expression (JAFPE) database.
- [11] Jolliffe, Ian . (2002). Principal Component Analysis. *Wiley Online Library*.
- [12] Lopes, André Teixeira., Aguiar, Edilson de., De Souza, Alberto F., Oliveira-Santos, Thiago . (2017). Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recognition*, 61, 610–628, 2017.
- [13] Lucey, Patrick., Cohn, Jeffrey F., Kanade, Takeo., Saragih, Jason., Ambadar, Zara., Matthews, Iain. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. *In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, p 94–101. *IEEE*.
- [14] Ma, Liying ., Khorasani,Khashayar. (2004). Facial expression recognition using constructive feedforward neural networks. Systems, Man, and Cybernetics, Part B: Cybernetics, *IEEE Transactions on*, 34 (3)1588–1595, 2004.
- [15] Mayer, Christoph., Eggers, Martin., Radig, Bernd (2014). Cross-database evaluation for facial expression recognition. *Pattern Recognition and Image Analysis*, 24 (1) 124–132, Jan 2014.
- [16] Munir, Asim., Hussain, Ayyaz., Khan, Sajid Ali., Nadeem, Muhammad., Arshid, Sadia (2018). Illumination invariant facial expression recognition using selected merged binary patterns for real world images. *Optik*, 158:1016–1025.
- [17] Nigam, Swati., Singh, Rajiv., Misra, AK. (20118). Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. *Multimedia Tools and Applications*, p. 1–23.
- [18] Norvig, Peter., Russell, Stuart. (2009). Artificial Intelligence: A Modern Approach. Pearson.
- [19] OpenCV. (1997). Open source computer vision library.
- [20] Owusu, Ebenezer., Zhan, Yongzhao., Mao, Qi Rong,. (2014). A neural-adaboost based facial expression recognition system. *Expert Systems with Applications*, 41 (7) 3383–3390.
- [21] John Platt., et al. (1999). Fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods: support vector learning*, 3, .
- [22] Shan, Caifeng, Gong, Shaogang., McOwan, Peter W (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27 (6) 803–816.
- [23] Shen, LinLin., Bai, Li., Fairhurst, Michael (2007). Gabor wavelets and general discriminant analysis for face identification and verification. *Image and Vision Computing*, 25(5) 553–563.
- [24] Sönmez, Elena Battini., Cangelosi, Angelo. (2017). Convolutional neural networks with balanced batches for facial expressions recognition. In Ninth International Conference on Machine Vision (ICMV 2016), 10341, p. 103410J. *International Society for Optics and Photonics*, 2017.
- [25] Viola, Paul., Jones, Michael. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. *In: Proceedings of the 2001 IEEE Computer Society Conference on*, 1, p. 1–511. *IEEE*, 2001.
- [26] Webb, A. R., Copsey, K. D. (2014). Statical Pattern Recognition. Wiley, West Sussex, UK, 2011.
- [27] WEKA. Weka 3: Data mining software in java. 2014.
- [28] YALE. Yale face database. 2000.
- [29] Zeng, Nianyin., Zhang, Hong., Song, Baoye., Liu, Weibo., Li, Yurong., Dobaie, Abdullah M. (2018). Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 273, 643–649.