Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 17, Iss. 2, pp. 4–14, 2019

A FACE RECOGNITION FRAMEWORK USING SELF-ADAPTIVE DIFFERENTIAL EVOLUTION

Guilherme Felippe Plichoski
Santa Catarina State University, Joinville, SC, Brazil
gplichoski@gmail.com

Chidambaram Chidambaram
Santa Catarina State University, Joinville, SC, Brazil
chidambaram@udesc.br

Rafael Stubs Parpinelli
Graduate Program in Applied Computing
Santa Catarina State University, Joinville, SC, Brazil
rafael.parpinelli@udesc.br

Abstract – It is well known that the development of face recognition (FR) systems is challenging under uncontrolled conditions often related to the variation of pose, illumination, expression, and occlusion. Also, to collect the necessary amount of images is hard to guarantee in many situations, e.g. ID cards, drivers licenses or visas, leading to the one sample per person (OSPP) problem. This work addresses the OSPP problem along with illumination and pose variation using an FR framework composed of a self-adaptive Differential Evolution, named FRjDE. The main feature of the framework stands on the use of the optimization algorithm for choosing which preprocessing and feature extraction strategies to use, as well as tunning their parameters. Also, by using the jDE algorithm, F and CR control parameters are self-adapted. Experiments are made using two well-known databases, named CMU-PIE and FERET. Results obtained from the FRjDE approach are compared against the FR framework using the standard DE algorithm and against results found in the literature. Results suggest that the proposed approach is highly competitive and well suited for face recognition.

Keywords – Illumination variation, pose variation, continuous optimization, evolutionary algorithms, machine learning, parameter control.

1. INTRODUCTION

Over the past decades, Face Recognition (FR) topic has brought the attention of the research community since traditional biometrics approaches became insecure with the evolution of technology. Also, FR systems have a distinct feature compared to other biometric systems since it does not require direct interaction with the user, allowing its application on environments with a high flow of people. The wide range of applications for FR, such as law enforcement, subjects identification, video surveillance, access control, social identification and criminal investigation led to several works proposing solutions to problems that arise on image acquisition [1]. Images acquired are often affected by imaging noises such as variation of pose, illumination, expression, and the use of adornments, which might lead to misclassification [2]. Among these, uncontrolled illumination is a critical factor for real-world applications due to the non-ideal imaging environments frequently encountered in practical situations [3]. In addition, pose variation affects classification in a way that the image loses spatial information, becoming necessary the development of approaches that deal with this condition. Despite these problems, to collect the necessary amount of data is hard to guarantee in many situations e.g. ID cards, drivers licenses or visas leading to the one sample per person (OSPP) problem [4]. Hence, to extract discriminative and robust features becomes a real challenge even to state-of-art approaches.

There are some recent works which attempt to solve the OSPP problem along with issues from image acquisition. Among them, there are methods employing sparse representation classifier [1] [4] [5], two-layer neural local-to-global feature learning framework [6], supervised locality preserving multimanifold (SLPMM) [7], single hidden layer analytic Gabor feedforward neural network (AGFN) [2] and multiple feature subspaces analysis (MFSA) [8]. Despite these, an increasing number of bio-inspired FR systems had emerged due to their intelligent problem-solving ability, scalability, flexibility, and adaptive nature [9].

In [10], an FR framework is presented and features the ability to self-adapt to conditions inherent in the database's images. The intent of the framework is to provide techniques for preprocessing and for feature extraction in which the optimization process, carried out by the optimization algorithm, gives as output the best set of techniques and parameters for the task. In this work, the authors compared the performance of a standard Particle Swarm Optimization algorithm and a standard Differential Evolution (DE) algorithm, recommending the use of DE algorithm in the framework.

However, it is well known that fixed parameters are not robust to all optimization problems and to tune these parameters is not trivial [11]. Hence, this paper is an extension of the work presented in BRACIS 2018 [10] and differs from previous work

in two points: i) we employ the FR framework using a self-adaptive version of DE [12], called FRjDE, in order to control the parameters of mutation factor (F) and crossover probability (CR). The goal is to minimize the number of parameters to adjust and analyze the performance of the framework; and ii) we used two well-known databases i.e. CMU-PIE [13] and FERET [14] considering the OSPP problem along with illumination and pose variation.

The remaining sections of this work are organized as follows: Section 2 presents the theoretical foundation and related work; Section 3 presents the FR framework; Section 4 shows the experimental methodology and datasets; Section V shows the results and analysis; and Section 6 presents the conclusions and some future directions.

2 BACKGROUND

In this section, we present the techniques employed for preprocessing, feature extraction and classification. These techniques serves as input for the proposed framework. It is worth to point out that they were chosen based on their performance in previous works and its popularity in literature specially along with bio-inspired algorithm [9].

2.1 PREPROCESSING TECHNIQUES

Some preprocessing techniques are available in the framework which are described next. The first technique available is called wavelet-based illumination normalization (WBIN) that is based on the discrete wavelet transform (DWT) [15]. In DWT the image is represented in terms of translations and dilatation of a scaling function and a wavelet function using a 2D filter bank consisting of low-pass and high-pass filters [15]. The different wavelet families make different trade-offs between how compact the basis functions are localized in space and how smooth they are. They can be biorthogonal (bior), Coiflets (coif), Daubechies (db), Haar, reversed biorthogonal (rbio), and Symlets (sym). Each family contains several variations according to its smoothness. The 2D-DWT produces four components, which are approximation (LL), horizontal (HL), vertical (LH), and diagonal (HH).

For the illumination normalization procedure, histogram equalization is performed in the approximation coefficients (LL) to achieve contrast enhancement. The fine details in the image can be controlled multiplying each element of the detail coefficient matrices (HL, LH and HH) with a scale factor (*Sf*). Then, the enhanced image is reconstructed using the inverse wavelet transform. The considered parameters within this technique are: the wavelet family (wf) that can assume 105 possible configurations and the scaling factor (sf) that can assume any value greater than 0.

The GIC [16], also known as power-law transformation, can control the overall brightness of an image by changing the parameter γ of its formula [16]. The only parameter to adjust within this technique is the parameter γ. If it assumes values lower than 1 will shift the image towards the darker end of the spectrum while γ values greater than 1 will make the image brighter.

The *Gaussian Blur* uses a 2D convolution operator with a Gaussian distribution to blur and remove details and noises from the image [17]. The parameters within this technique are: Gaussian kernel size, which must assume positive and odd values, and Gaussian kernel standard deviation in X (sdX) and Y (sdY) directions [1].

An edge is a set of pixels in the image where intensity abrupt changes, representing the shape of the face, eyes, nose and other discriminant features. In [18], a review of edge detection techniques was performed which featured the *Laplacian* with better results for visual perception and edge quantity. The parameters that have been considered in this technique are: the aperture size used to compute the second-derivative filters which must be positive and odd, a scale factor (sf) for the computed Laplacian values and a delta value that is added to the results prior to storing them in the resulting image [1].

2.2 FEATURE EXTRACTION TECHNIQUES

In this section, some feature extraction techniques are described. The first feature extraction technique included in the FR framework is the *2D-DWT*, which was previously introduced. According to [16], the information in the approximation component has the most discriminant features. Hence, only this sub-band was used. The considered parameters within this technique are: the wavelet family (wf) and the number of levels of decomposition (n).

The discrete cosine transform (DCT) was also added to the framework. The DCT technique is popular for image processing because of its energy compaction property. It transforms the image from spatial to frequency domain using cosine functions. In frequency domain, the relevant features tend to be concentrated in low-frequency components or at the corner of the spectrum [19].

According to [20], the local binary patterns (LBP) texture descriptor has some advantages that are: easy to implement, invariant to monotonic illumination changes, and low computational complexity. Its canonical version is used in this work from which the features are extracted directly from the input image using a 3x3 neighborhood. The neighborhood structure is a set of pixels taken from a square neighborhood which are compared against the value of the central pixel (thresholded) resulting in an 8 bits vector converted to decimal. Then, the resultant value is used to represent the pixel.

2.3 CLASSIFICATION

For the classification task, each image to be classified is compared to each image in the database in terms of similarity returning the most similar one. Then, a match occurs if both images are from the same person. This classification is carried out by the well known k-Nearest Neighbors (k-NN) classifier, where $k = 1$ [21]. As k-NN is a common classifier in the literature, it

[1] *The OpenCV Reference Manual*, third edition, February 2018. Retrieved from https://docs.opencv.org/

suggests a reliable choice for this case study. In this work, four different distance measures are available in the framework. The first distance measurement implemented is the Manhattan distance (L_1-norm), that is the distance between two points alongside the axis at right angles. Similarly, the well-known Euclidean distance (L_2-norm) measures the distance between two points directly and was also employed. In addition to these, the L_{2SQR}-norm (the square of the L_2-norm) and L_∞-norm were also implemented.

2.4 DIFFERENTIAL EVOLUTION AND SELF-ADAPTIVE DIFFERENTIAL EVOLUTION (JDE)

The Differential Evolution (DE) algorithm is a metaheuristic proposed by Storn and Price in 1996 for continuous optimization problems [22]. It consists of a population of size NP of solution vectors of size D, where new individuals are generated by recombination of individuals of the previous generation. Each individual has a fitness value proportional to how good it is and this value is responsible for guiding the DE in the search space. A newer individual compete with the previous one and must have a better fitness in order to replace the competing individual.

The input parameters that enable to control the search are the population size, the crossover probability (CR) and the mutation factor (F). DE has a nomenclature that describes the adopted configuration. This takes the form of $DE/a/b/c$, where a represents the solution vector to be perturbed. The most commonly used strategies are selecting at random (*rand*) and selecting the best solution (*best*). The b identifies the number of difference vectors used in the perturbation of a. Last, c is the recombination operator performed such as bin for binomial and exp for exponential [23]. The most popular configuration is demonstrated in the referenced pseudo-codes i.e. $DE/rand/1/bin$.

The standard DE keeps the control parameters fixed during the whole optimization process and are defined by the user. On the other hand, the jDE [12] uses a self-adaptive strategy to control F and CR parameters. Each individual i in the population is extended with two new dimensions representing F_i and CR_i control parameters which are adjusted on-the-fly. Better values of these control parameters lead to better individuals which, in turn, are more likely to survive and produce offspring and, hence, propagate these parameter values [12].

The parameters F_i^{G+1} and CR_i^{G+1} are calculated as shown in Equation 1 and Equation 2 where $rand_{\{1,\cdots,4\}}$ are uniform random values within the range $[0, 1]$, τ_1 and τ_2 represent fixed probabilities to adjust control parameters F and CR, respectively, and F_{lower} and F_{upper} are the fixed lower and upper bounds that the parameter F can assume.

$$F_i^{G+1} = \begin{cases} F_{lower} \ + \ rand_1 \ \cdot \ F_{upper} & \text{if } rand_2 < \tau_1, \\ F_i^G & \text{otherwise.} \end{cases} \tag{1}$$

To the new F_i is assigned a value ranging from $[F_{lower}, F_{upper}]$ and to the new CR_i is assigned a value from $[0, 1]$. F_i^{G+1} and CR_i^{G+1} are obtained before the mutation is performed. Hence, their new values influence the mutation, crossover and selection operations of the new solution vector $\vec{p}_i^{G+1}$.

$$CR_i^{G+1} =' \begin{cases} rand_3 & \text{if } rand_4 < \tau_2, \\ CR_i^G & \text{otherwise.} \end{cases} \tag{2}$$

2.5 RELATED WORK

The works described in this section were chosen based on the FR problem being tackled (illumination and pose variation) and the experimental methodology employed using the CMU-PIE and FERET databases. Also, the OSPP problem is a real-world condition and is considered in the related work.

In [1], a customized dictionary-based face recognition approach is proposed using the extended joint sparse representation. The proposed approach learn a customized variation dictionary from the on-location probing face images and then utilizes the information of both the customized dictionary and the gallery samples to classify the probe samples. [4] presents a model called Local Robust Sparse Representation (LRSR) to tackle the problem of query images with various intra-class variations. The idea is to combine the local sparse representation model with a patch-based generic variation dictionary learning model to predict the possible facial intra-class variation of the query images. The authors stated that their methodology outperforms the state-of-art approaches. In [5] is proposed an additive strategy to the Extended Sparse Representation Classifier (ESRC) aiming to solve the mutual influences of the reflectance and illumination in intra-class variant bases of the ESRC and remove the redundant identity information of the generic face. Two additive models were introduced: the Reflectance and Illumination (RandL) and the High-and Low-frequency (HandL). The final classification is determined by the sum of two weighting residuals. In their experiments, the Wavelet Denoising Model (WDM) and the Logarithmic Total Variation (LTV) are employed to extract facial features in R&L_ESRC and H&L_ESRC, respectively. [6] proposed a two-layer neural local-to-global feature learning framework to address OSPP face recognition. In the first layer, the objective-oriented local features are learned by a patch-based fuzzy rough set feature selection strategy and in the second layer, global structural information is extracted from local features by a sparse auto-encoder. The authors claimed that their framework can achieve superior performance than other state-of-the-art feature learning algorithms for OSPP face recognition. In [7] the authors presented a supervised learning method called supervised locality preserving multimanifold (SLPMM). In their approach, two graphs are made to represent the information inside every manifold and among different manifolds, with this it simultaneously maximizes the between-manifold scatter and minimizes the within-manifold scatter. [2] proposed a single hidden layer analytic Gabor feedforward neural network (AGFN).

The input layer receives raw images and propagates geometrically localized sub-image patches to the hidden layer, then vertically and horizontally accumulated Gabor magnitude features are extracted over several orientation and scale settings. Moreover, a whitening process with dimension reduction is employed and the extracted features are fused or classification decision. Finally, [8] present a multiple feature subspaces analysis (MFSA) approach, which takes advantage of facial symmetry. They divide each enrolled face into two halves about the bilateral symmetry axis and further partition every half into several local face patches. Then, they cluster all the patches into multiple groups according to their locations at the half face and learn a feature subspace for each group. The classification is performed employing the k-NN classifier.

Most methods present in literature are limited because they are designed to solve a specific set of image conditions, lacking the ability to adapt themselves to different conditions. Real-world applications require robustness, since new challenges arises regarding images conditions. Following, we present the proposed framework which aims to fill that gap in existing methods.

3 THE FR FRAMEWORK

The FR framework provides techniques for preprocessing and feature extraction in which the optimization process, carried out by the bio-inspired algorithm, gives as output the best set of techniques for the classification task [10]. Besides that, the parameters of selected techniques are also optimized by the framework as well as the distance metric employed by the classifier.

Figure 1 represents the FR system flowchart. The dataset should be split into three subsets: training, validation and test. In the training stage, the validation subset is used to evaluate the candidate solutions generated by the optimization algorithm based on the training subset. In the test stage, the test subset is used to evaluate the optimized model generated after training. In both the training and test stages, the fitness evaluation is done based on the recognition rate.

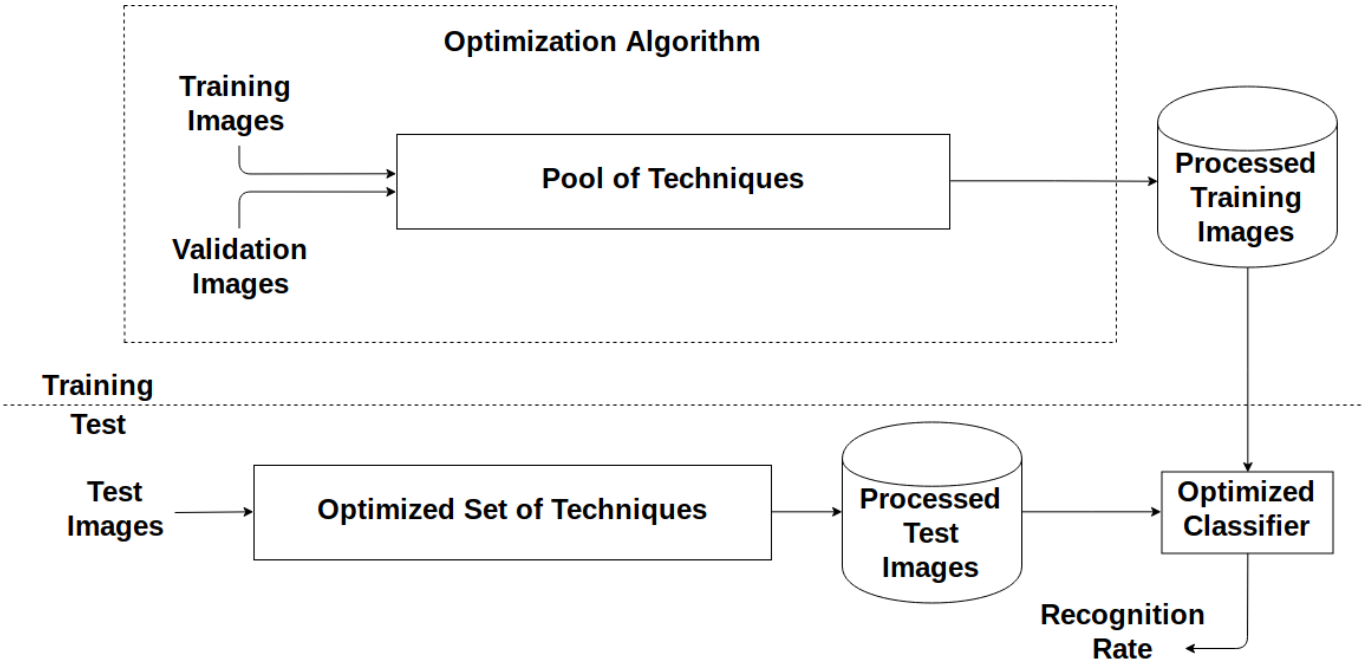


Figure 1: FR system flowchart.

The techniques introduced in Section 2.1 and 2.2 are available in the framework, as well as the distance measures presented in Section 2.3. For preprocessing, the first technique employed was WBIN to adjust contrast and enhance face structure. After, the GIC for brightness adjustment and Gaussian Blur to remove noise from the image were added to the pool. Last, a technique for edge detection, the Laplacian was included in our framework. It is worth to point out that the sequence in which these techniques are applied are in the same order as presented here. Hence, in this implementation, the optimization algorithm can choose if the techniques will be applied and its parameters but not the order they will be applied. Considering feature extraction, the order of the techniques employed does not affect the results since the features are concatenated. Despite the techniques implemented in this instance of the framework, any preprocessing or feature extraction technique may be implemented in the pool of techniques. Figure 2 illustrates the employed techniques and their respective parameters.

The optimization algorithm acts in the training stage and the techniques, as well as their parameters, are encoded in the solution vector as real-coded variables. Figures 3 shows how the solution vector is encoded, consisting of WBIN, GIC, Gaussian Blur, Laplacian Edge Detection, DWT, DCT, LBP, and the distance metric of k-NN, respectively. The parameter $flag$ represents if the technique will be used or not. As the optimization algorithms operate in the continuous domain, the parameter $flag$ is discretized to 0 or 1. Also, due to the stochastic nature of the optimization process, if a parameter assumes a value that is out of its bounds, the value is set to the closest bound. Any real-coded algorithm for continuous optimization may be used in the FR framework.

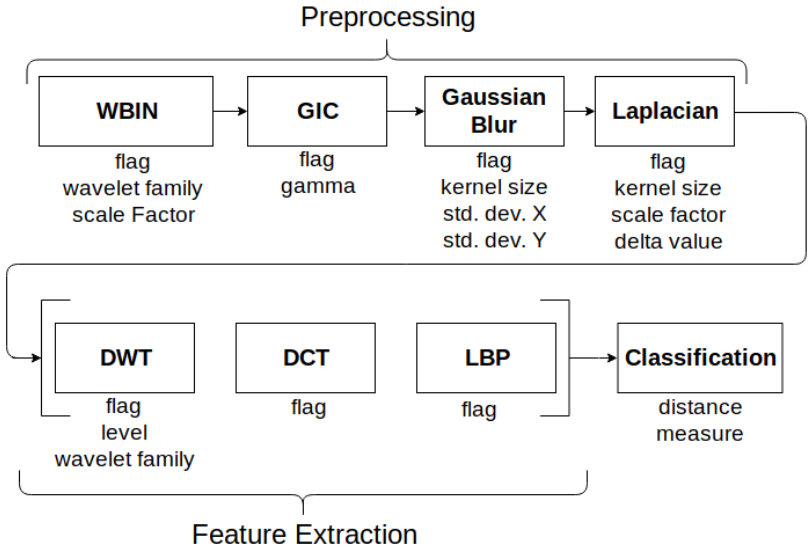


Figure 2: Pool of techniques employed in the FR flowchart.

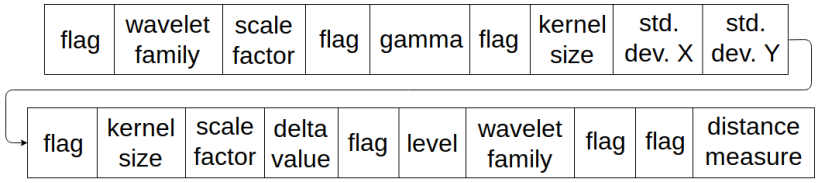


Figure 3: Solution encoding.

The optimization algorithm starts by initializing the population with random solution vectors and, as the iterations cycle begins, the solutions are modified accordingly to the algorithm strategy. A predefined number of iterations is established as a stopping criterion. When the stop criterion is reached, the setting that achieved the best recognition rate is selected as the optimized solution. To test the classification model, the test set is transformed using the optimized set of techniques and thus, generating the processed test images dataset. Finally, the processed test images are classified by the optimized k-NN classifier using the processed training dataset as reference.

4 EXPERIMENTS

As DE and jDE are stochastic algorithms, multiple runs are required. The measurements needed for each algorithm in each dataset was established using a confidence interval of 95% with a standard deviation of 2.5% [24]. Based on this criterion, 10 runs were performed for each algorithm in each dataset. The recognition rate, which is the ratio between the number of reference face images assigned to the correct class and the total amount of reference face images, and processing time are used as performance metrics.

Table 1 shows the parameters for DE and the common parameters for both De and jDE algorithms. All parameters were set using literature recommendations [23].

Table 1: Employed parameters for both algorithms

Parameters	
DE parameters	
F	0.8
CR	0.9
Common parameters	
$Population\ Size$	20
$Stop\ Criteria$	100 iterations

Statistical tests are required to ensure statistical significance in the results obtained. Based on this, we first applied the Shapiro and Wilk normality test [25] and, once verified, the Wilcoxon signed-rank paired test [26] is applied. Both tests were performed under a 95% confidence interval.

To better understand the behavior of the optimization algorithms, convergence and diversity charts are generated. The convergence chart represents the recognition rate of the best solution at each iteration. The population's diversity chart is obtained using Equation 6 from Corriveu et al. [27]. This chart represents how spread the solutions are in the search space, this way, it is possible to draw whether the algorithm is performing a more local or global search at each iteration.

All experiments were run in an Intel Core i7-4770 computer with 16 GB RAM memory using Python programming language with multiprocessing library for parallelism. Parallelism is achieved distributing the fitness evaluation of optimization algorithms population in 7 out of 8 available processing cores. Following, we describe the datasets used in this work.

4.1 CMU-PIE DATABASE

The CMU-PIE [13] database consists of 41,368 images of 68 subjects. Each subject has several images under 13 different poses and 43 illumination conditions, and with 4 expressions. For experimentation, it was selected five subsets based on the camera used (i.e. C05, C07, C09, C27 and C29). Each camera is set in different angles, this way presenting five different poses: 1) looking right; 2) up; 3) down; 4) front; and 5) left, respectively. Figure 4 illustrates the five subsets. For a fair comparison with literature, we have selected about 24 images per subject with different illumination conditions according to database documentation. Each one of the five subsets is evaluated independently and grouped together (named C_{All}). When using the subsets individually, we aim to analyze the OSPP problem, in which only one single image per person is available in the training set. That being said, the first image of every subject from each subset is used. The remaining images are randomly split in half representing the validation and test sets. In the experiment using C_{All}, the validation, test and training sets from each subset are grouped together, respectively. This experiment makes possible to evaluate the proposed approach in a set containing intra-pose variation. It is worth to point out that this last experiment does not characterize the OSPP problem.

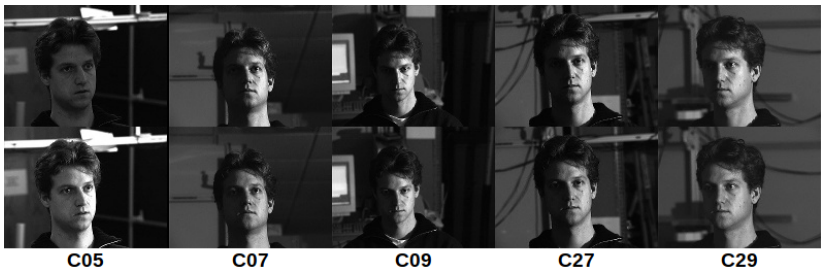


Figure 4: Example of CMU-PIE database subsets C05, C07, C09, C27 and C29, respectively.

4.2 FERET DATABASE

The b-series subset from FERET database [14] was employed, which consists of 1,800 images from 200 subjects with facial pose angles ranging from +60^o to -60^o and size of 256x384. Table 2 presents the description of subsets used in this experiment and Figure 5 shows some face images samples. For comparison purposes with the literature, the frontal subset *ba* is used as training set, and all the remaining subsets are grouped together as validation and test sets (b_{All}). Here we consider the OSPP problem as well.

Table 2: Five subsets according to the light angle source directions

	bb	*bc*	*bd*	*be*	*ba*	*bf*	*bg*	*bh*	*bi*
Pose angle	+60	+40	+25	+15	0	-15	-25	-40	-60
Number of images per subject	1	1	1	1	1	1	1	1	1
Total images	200	200	200	200	200	200	200	200	200

Figure 5: Example of FERET b-series subsets.

5 RESULTS AND ANALYSIS

This section shows and analyses the results obtained by the FR framework using the self-adaptive version of DE, namely FRjDE. Results are compared with its standard version FR-DE and with results found in the literature.

Table 3 presents the results obtained using FRjDE and FR-DE algorithms on each CMU-PIE subset. The last line shows the average recognition rate. Regarding the test stage, it is possible to realize that a gain in accuracy is achieved when jDE is employed for each subset except for C_{All}. Considering the statistical tests, only the subsets C09 ($p - value = 0.0113$) and C27 ($p - value = 0.0047$) presented significant differences. Despite that, it is possible to observe an average gain in accuracy in favor of jDE of roughly 1%. In the training stage, no statistical difference was verified. Table 4 shows the recognition rates for FRjDE and FR-DE on FERET database. For this dataset, statistical difference in favor of jDE ($p - value = 0.0069$) was verified in the test stage of 3.24% better on average.

In a general analysis, it is possible to conclude that adjusting the control parameters during the optimization process with jDE leads to a more precise and accurate search for an optimized solution.

Table 3: Recognition rates (%) using FRjDE and FR-DE on CMU-PIE subsets

	FRjDE		**FR-DE**	
	Training	**Test**	**Training**	**Test**
C05	100	99.96 ± 0.08	100	99.95 ± 0.16
C07	100	99.52 ± 0.10	99.78 ± 0.36	99.37 ± 0.34
C09	100	99.97 ± 0.06	100	99.22 ± 1.61
C27	96.51 ± 0.53	96.40 ± 0.28	96.90 ± 1.52	95.07 ± 1.28
C29	100	99.86 ± 0.12	100	99.75 ± 0.25
C_{All}	99.86 ± 0.04	99.54 ± 0.26	99.88 ± 0.09	99.74 ± 0.10
Average	99.39 ± 1.41	99.20 ± 1.39	99.42 ± 1.24	98.85 ± 1.87

Table 4: Recognition rates (%) using FRjDE and FR-DE on FERET dataset

	FRjDE		**FR-DE**	
	Training	**Test**	**Training**	**Test**
FERET	80.56 ± 1.70	79.53 ± 0.22	80.75 ± 2.50	77.51 ± 2.26

Regarding processing time, Table 5 presents the average running times for each CMU-PIE subset and for the FERET database, respectively. From Table 5, no significant difference can be verified comparing FRjDE and FR-DE. However, comparing the absolute values obtained, it is possible to observe that FRjDE was faster to train and test in general. This is possibly a reflection of a better search in the problem space due to the self-adjustment of the control parameters in FRjDE.

Table 5: The average processing time in training and test stages on CMU-PIE subsets and FERET dataset

	FRjDE		**FR-DE**	
	Training	**Test per face image**	**Training**	**Test per face image**
C05	00h 02m 7s ± 00h 00m 23s	0.0724s ± 0.0362s	0h 1m 43s ± 0h 0m 17s	0.0760s ± 0.0234s
C07	0h 4m 11s ± 00h 03m 58s	0.0310s ± 0.0440s	0h 34m 46s ± 0h 41m 15s	0.0416s ± 0.0399s
C09	0h 03m 15s ± 0h 2m 0s	0.0653s ± 0.0424s	0h 3m 20s ± 0h 3m 2s	0.0588s ± 0.0375s
C27	2h 4m 47s ± 0h 42m 27s	0.0097s ± 0.0024s	2h 23m 3s ± 1h 5m 42s	0.0571s ± 0.0775s
C29	0h 2m 18s ± 0h 01m 8s	0.0898s ± 0.0008s	0h 2m 14s ± 0h 0m 27s	0.0735s ± 0.0245s
C_{All}	6h 3m 25s ± 5h 45m 54s	0.1132s ± 0.0125s	8h 11m 46s ± 6h 16m 45s	0.1043s ± 0.0059s
FERET	2h 13m 37s ± 0h 42m 11s	0.0093s ± 0.0025s	3h 25m 15s ± 1h 17m 29s	0.0152s ± 0.0069s

Tables 6 and 7 show the settings found by FRjDE in training step for C_{All} CMU-PIE subset and FERET database, respectively. Each one of the ten executions is shown in rows, while the techniques chosen are in columns. A dash symbol means that the technique was not employed in the related execution. Since similar conclusions can be drawn from the CMU-PIE subsets, we only present the settings for C_{All} subset. Considering C_{All} subset, it is possible to realize that WBIN was chosen in every execution except one, and the LBP technique was always chosen. This demonstrates the importance of these techniques to deal with the conditions on CMU-PIE database. Also, the preferred distance measure was the Manhattan distance (L_1). On the other hand, the Gaussian Blur, Laplacian and DWT were never chosen. Differently from CMU-PIE database, the Gaussian Blur and DWT was chosen in every execution for FERET database. Also, we can observe that the Manhattan distance was once more the preferred distance measure.

The best results achieved by the FR framework using jDE in the two databases are compared with the results obtained by related work found in the literature (Section 2.5). The best average result obtained by FRjDE in training and test stages was

Table 6: Settings found by FRjDE from training step for C_{All} CMU-PIE subset

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	bior4.4	0.8726	1.5110	-	-	-	-	-	-	-	-	Yes	Yes	L_1
2	db11	0.2381	-	-	-	-	-	-	-	-	-	-	Yes	L_{2sqr}
3	rbio3.5	0.2715	-	-	-	-	-	-	-	-	-	-	Yes	L_1
4	coif14	0.0000	-	-	-	-	-	-	-	-	-	-	Yes	L_2
5	db6	0.2487	1.0566	-	-	-	-	-	-	-	-	-	Yes	L_2
6	sym20	0.8519	2.2367	-	-	-	-	-	-	-	-	-	Yes	L_1
7	db27	0.0000	-	-	-	-	-	-	-	-	-	-	Yes	L_1
8	-	-	2.3445	-	-	-	-	-	-	-	-	Yes	Yes	L_1
9	coif7	0.0000	3.0000	-	-	-	-	-	-	-	-	Yes	Yes	L_1
10	-	-	2.8576	3	2.1826	5.0000	3	1.6230	1.9506	-	-	-	Yes	L_{2sqr}

Table 7: Settings found by FRjDE from training step on FERET dataset

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	-	-	1.5512	13	1.8979	3.9317	-	-	-	rbio2.2	4	Yes	-	L_1
2	-	-	-	23	5.0000	4.0797	-	-	-	bior5.5	3	Yes	-	L_1
3	-	-	-	23	4.3469	2.6568	-	-	-	sym4	3	-	-	L_1
4	-	-	1.3597	31	4.4828	3.9360	-	-	-	sym4	3	-	-	L_1
5	-	-	-	21	4.9787	4.2976	-	-	-	bior6.8	2	Yes	-	L_1
6	-	-	-	19	4.0332	1.0000	-	-	-	rbio1.5	3	Yes	-	L_1
7	-	-	-	5	4.2975	2.4174	-	-	-	bior5.5	3	Yes	-	L_1
8	-	-	-	31	5.0000	1.2847	-	-	-	sym10	2	-	-	L_1
9	-	-	-	3	2.1067	1.0000	-	-	-	rbio3.3	3	-	-	L_1
10	-	-	1.2344	11	4.9958	1.0161	-	-	-	sym15	2	-	-	L_1

chosen for comparison. Table 8 shows the recognition rates achieved by FRjDE and related works on CMU-PIE database. The last line shows the results obtained by using the most common parameters found in the literature accomplished with the set of techniques defined by our approach. It is worth to remember that all these works consider the OSPP problem. As one can see, the proposed approach presents outstanding accuracy when compared to the related work considering these subsets. Also, the experiment considering pose variation within all subsets (C_{All}) shows that the proposed approach achieves high recognition rates even when slightly pose variation occurs. Overall, considering illumination compensation properties, the presented method proved to be competitive with related works for this dataset.

From Table 8, regarding the results achieved using literature settings, as different techniques reached the same highest recognition rate, the criterion used to selected the experiment to try out with literature settings was less processing time. For the C05 subset, experiment 4 was chosen where DWT, DCT, GIC and Laplacian techniques were employed with Manhattan distance. So, we set scale factor to 3 and the haar wavelet for DWT, the γ parameter from GIC to 2 and the kernel size to 3x3 with no scale or delta value for Laplacian using the Euclidean distance. As one can see, it reached also 100% recognition rate showing that the parameters are insensitive for this case, this way once more proving the easiness of this subset. On the other hand, the parameters for experiment 7 using C07 dataset showed to be sensitive dropping 10.52% recognition rate when using literature settings. In this experiment, the GIC and Laplacian technique were employed with Manhattan distance. We set the same values used in subset C05 with Euclidean distance. In experiment 5 from C09, the techniques chosen by jDE algorithm were DWT, GIC and Laplacian with Manhattan distance. Employing literature parameters, a slightly drop was observed (≈4.09%). Considering experiment 9 from subset C27, a higher drop in accuracy was verified (≈19.48%). In this, Gaussian Blur and WBIN were employed besides GIC and DCT using the Manhattan distance. For Gaussian Blur, the kernel size was set to 3x3 and the standard deviations were computed from kernel size, and for WBIN, the haar wavelet was employed with scale factor 3. Experiment 2 using jDE was chosen to comparison using C29 subset. A considerable accuracy drop of ≈23.78% was verified. The techniques employed in this experiment were DCT and WBIN with Manhattan distance. Finally, experiment 9 from C_{All} revealed an accuracy drop of ≈17.28%, once more proving the importance of parameters adjustment. The techniques employed were WBIN and LBP.

Table 8: CMU-PIE database recognition rates (%) of FRjDE compared to related works

Method	C05	C07	C09	C27	C29	C_{All}
CD-EJSR [1]	50	-	-	-	-	-
LRSR [4]	-	93.46	93.63	-	-	-
R&L_ESRC [5]	-	-	-	≈92.00	-	-
TLFL [6]	-	54.87	56.47	-	49.71	-
Our method	**100**	**99.67**	**100**	**96.85**	**100**	**99.80**
Literature settings	100	89.15	95.91	77.37	76.22	82.52

Table 9 presents the results obtained by the related work and by the FR framework using jDE on FERET database. Partial results of each subset and the average (b_{All}) are shown in columns, while the algorithms are shown in rows. The last row shows results achieved using the techniques found by the framework but with parameter values commonly used in literature. For

comparison, the setup obtained by experiment 3 was chosen. Through this, one may realize that even though our approach is not specific to handle pose variation, competitive results were achieved considering the works presented. Also, when analyzing results obtained by the SLPMM and the MFSA+ approaches, the superiority of the results obtained by the proposed approach are evident. Furthermore, our approach achieved overall best average using b_{All} subset compared to all algorithms highlighting the robustness of the proposed method, since b_{All} considers all different conditions contained in FERET b-series dataset. On the other hand, for most subsets (*bd*, *be*, *bf*, *bg*, and *bh*) the method from [2] achieved better results. This was expected since their work is focused on a pose-invariant approach. From this dataset, the greater the variation of pose angle, the more difficult it is to solve by an FR system. Hence, our approach achieved the best performance in the subsets with greater pose variation, i.e. *bb*, *bc* and *bi*, and also was very competitive in the *bh* subset. Once again, it is possible to draw from the results using literature settings the importance of tuning the parameters of the employed techniques.

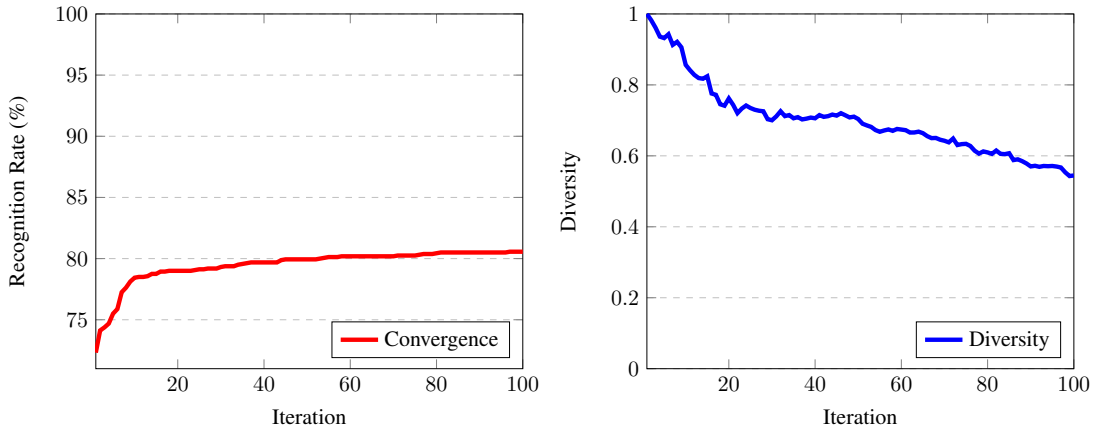
Table 9: FERET database recognition rates (%) of FRjDE compared to related works

Method	*bb*	*bc*	*bd*	*be*	*bf*	*bg*	*bh*	*bi*	b_{All}
frequency SLPMM [7]	15.5	19.5	43.5	77	82	51.5	28.5	14.5	41.5
AGFN$_{hor}$ [2]	45.4	60.2	**93.6**	**97.6**	**99.3**	**96.7**	**78.2**	52.4	77.92
MFSA+ [8]	-	-	61	76.5	85	52	-	-	-
Our method	**67.2**	**82.7**	89.4	90.5	88.8	87.2	75	**59.4**	**80**
Literature settings	42.2	66.6	73.8	80.5	80	66.6	55.5	35.5	62.6

*Best results are highlighted in bold.

Figure 6 shows the convergence and diversity charts for FERET database using FRjDE. Only the charts of FERET dataset are presented because the same behavior of convergence and diversity is obtained from CMU-PIE subsets. In the convergence chart, it is possible to observe that close to the 40th iteration the best solution improves at a slow rate but continuously improving when the stop criterion is reached. Due to the multimodal search space nature of the problem, the diversity in our experiments remained high during all optimization step.

Figure 6: Convergence and Diversity charts, respectively, on FERET database



6 CONCLUSION AND FUTURE WORK

To the best of our knowledge, from early studies to the most recent ones, it can be said that there is no generic technique for FR that is immune to the challenges that arise when a face is captured in uncontrolled conditions. Hence, we employed an FR framework with the aid of a self-adaptive DE, namely FRjDE. The main aspect is the definition of which set of techniques and its ideal parameters values should be used for the FR using jDE. In this work, we implemented several illumination compensation techniques to handle the problem of illumination variation, as it is a critical factor. The well known CMU-PIE and FERET face datasets were used to evaluate the proposed approach since it permits to compare the results with relevant works found in the literature.

Considering CMU-PIE database, both algorithms showed to be suitable in compensation illumination variation achieving more than 99% of recognition rate in most subsets. Also, 100% was reached on subsets C05, C09 and C29. Besides that, in the experiment with C$_{All}$ subset, it was possible to verify the capability of the proposed approach in dealing with slight changes of pose. For the FERET database, the recognition rates do not reach real-world levels of accuracy, however, the results achieved are outstanding since the dataset contains pose variation. Regarding the usage of self-adaptive control parameters, it proved to improve recognition rate significantly in most cases. Hence, we can state that employing jDE instead of standard DE lead to higher accuracy considering the dataset conditions. It is worth to point out that the OSPP problem was considered for both

datasets. Besides these conclusions, it can be observed that the present FR problem is highly multimodal since different sets of techniques can achieve similar (or the same) results.

In this work, we only consider a specific sequence of the employed preprocessing techniques. We suggest to include the possibility of interchange these techniques in the optimization step. Despite the fact that high recognition rates were achieved, it seems that there is still room for improvement. With this, one could use some strategy such as meta-learning to fine tune the previously achieved solution. To achieve robustness in relation to the problems that arise on image acquisition, different kinds of problems need to be tackled. For this, other techniques must be added to the framework and other datasets should be tested. Also, a physical interpretation analysis of the images could be performed considering the parameters found for the preprocessing techniques. Finally, the testing phase is performed in the range of milliseconds, however, the training phase still lasts hours for some datasets. This does not invalidate our approach for real-world application, but some parallel programming strategies or massively parallel hardware architectures, such as GPGPUs, could be used.

REFERENCES

[1] K. Shang, Z.-H. Huang, W. Liu and Z.-M. Li. "A single gallery-based face recognition using extended joint sparse representation". *Applied Mathematics and Computation*, vol. 320.

[2] B.-S. Oh, K.-A. Toh, A. B. J. Teoh and Z. Lin. "An Analytic Gabor Feedforward Network for Single-Sample and Pose-Invariant Face Recognition". *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2791–2805, 2018.

[3] W. Zhang, X. Zhao, J.-M. Morvan and L. Chen. "Improving shadow suppression for illumination robust face recognition". *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 3, pp. 611–624, 2019.

[4] J. Gu, H. Hu and H. Li. "Local robust sparse representation for face recognition with single sample per person". *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 2, pp. 547–554, 2018.

[5] C. Hu, X. Lu, M. Ye, W. Zeng and Y. Du. "Illumination robust single sample face recognition based on ESRC". *Multimedia Tools and Applications*, vol. 76, no. 24, pp. 26523–26550, 2017.

[6] Y. Guo, L. Jiao, S. Wang, S. Wang and F. Liu. "Fuzzy sparse autoencoder framework for single image per person face recognition". *IEEE transactions on cybernetics*, vol. 48, no. 8, pp. 2402–2415, 2017.

[7] N. Mehrasa, A. Ali and M. Homayun. "A supervised multimanifold method with locality preserving for face recognition using single sample per person". *Journal of Central South University*, vol. 24, no. 12, pp. 2853–2861, 2017.

[8] Y. Chu, L. Zhao and T. Ahmad. "Multiple feature subspaces analysis for single sample per person face recognition". *The Visual Computer*, vol. 35, no. 2, pp. 239–256, 2019.

[9] G. F. Plichoski, R. S. Parpinelli and C. Chidambaram. "Swarm intelligence and evolutionary computation approaches for 2D face recognition: a systematic review". *Revista Brasileira de Computação Aplicada*, vol. 10, no. 2, pp. 2–17, 2018.

[10] G. Plichoski, C. Chidambaram and R. S. Parpinelli. "A Face Recognition Framework for Illumination Compensation Based on Bio-Inspired Algorithms". In *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 284–289. IEEE, 2018.

[11] R. S. Parpinelli, G. F. Plichoski, R. S. Da Silva and P. H. Narloch. "A Review of Techniques for On-line Control of Parameters in Swarm Intelligence and Evolutionary Computation Algorithms". *International Journal of Bio-Inspired Computation (IJBIC)*, vol. 13, no. 1, pp. 1–20, 2019.

[12] J. Brest, V. Zumer and M. S. Maucec. "Self-adaptive differential evolution algorithm in constrained real-parameter optimization". In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pp. 215–222. IEEE, 2006.

[13] T. Sim, S. Baker and M. Bsat. "The CMU pose, illumination, and expression (PIE) database". In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 53–58. IEEE, 2002.

[14] P. J. Phillips, H. Wechsler, J. Huang and P. J. Rauss. "The FERET database and evaluation procedure for face-recognition algorithms". *Image and vision computing*, vol. 16, no. 5, pp. 295–306, 1998.

[15] S. Du and R. Ward. "Wavelet-based illumination normalization for face recognition". In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, pp. II–954. IEEE, 2005.

[16] K. Manikantan, M. S. Shet, M. Patel and S. Ramachandran. "DWT-based illumination normalization and feature extraction for enhanced face recognition". *International Journal of Engineering & Technology*, vol. 1, no. 4, pp. 483–504, 2012.

[17] R. C. Gonzalez and R. E. Woods. *Digital Image Processing, Pretice Hall, 2008.* Pearson, London, 2017.

[18] A. Kushwah, K. Gupta, A. Agrawal, G. Jain and G. Agrawal. "A Review: Comparative Study of Edge Detection Techniques". *International Journal*, vol. 8, no. 5, pp. 2528–2531, 2017.

[19] S. Rao and M. B. Rao. "A novel triangular DCT feature extraction for enhanced face recognition". In *Intelligent Systems and Control (ISCO), 2016 10th International Conference on*, pp. 1–6. IEEE, 2016.

[20] L. Liu, P. Fieguth, Y. Guo, X. Wang and M. Pietikäinen. "Local binary features for texture classification: Taxonomy and experimental study". *Pattern Recognition*, vol. 62, no. C, pp. 135–160, 2017.

[21] T. Cover and P. Hart. "Nearest neighbor pattern classification". *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[22] R. Storn and K. Price. "Minimizing the real functions of the ICEC'96 contest by differential evolution". In *Proceedings of IEEE International Conference on Evolutionary Computation*, pp. 842–844. IEEE, 1996.

[23] J. Brownlee. *Clever Algorithms: Nature-Inspired Programming Recipes*. Lulu.com, Melbourne, first edition, 2011.

[24] D. J. Lilja. *Measuring computer performance: a practitioner's guide*, volume 1. Cambridge university press, Cambridge, Reino Unido, 2005.

[25] S. S. Shapiro and M. B. Wilk. "An analysis of variance test for normality". *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[26] F. Wilcoxon. "Individual comparisons by ranking methods". *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[27] G. Corriveau, R. Guilbault, A. Tahan and R. Sabourin. "Review of phenotypic diversity formulations for diagnostic tool". *Applied Soft Computing*, vol. 13, no. 1, pp. 9–26, 2013.