From current literature, it is well known that Face Recognition (FR) systems cannot perform well under uncontrolled conditions, thus there are no general and robust approaches with total immunity to all conditions. Hence, we present an adaptive FR framework with the aid of bio-inspired optimization algorithms. This approach implements several preprocessing and feature extraction techniques. The main feature of the present work stands on the use of the optimization algorithm which is responsible for choosing which strategies to use, as well as tuning their parameters. In this work, we employed three optimization algorithms for this task, namely Particle Swarm Optimization (PSO), Differential Evolution (DE) and the Self-adaptive Differential Evolution (jDE), aiming to address the illumination compensation problem. According to the proposed FR framework, the optimization algorithm can choose any set of the following techniques and tune their parameters: the Gamma Intensity Correction (GIC), the Wavelet-based Illumination Normalization (WBIN), the Gaussian Blur, the Laplacian for edge detection, the Discrete Wavelet Transform (DWT), the Discrete Cosine Transform (DCT), and the Local Binary Patterns (LBP). Experimental results are achieved using three well known databases i.e. The Yale Extended B, CMU-PIE and FERET, which confirmed that the proposed approach is suitable for FR.

Orientador: Rafael Stubs Parpinelli

Coorientador: Chidambaram Chidambaram

JOINVILLE, 2019

ANO 2019

GUILHERME FELIPPE PLICHOSKI |A FACE RECOGNITION FRAMEWORK BASED ON BIO-INSPIRED OPTIMIZATION ALGORITHMS

UDESC

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
PROGRAMA DE PÓS GRADUAÇÃO EM COMPUTAÇÃO APLICADA

DISSERTAÇÃO DE MESTRADO

A FACE RECOGNITION FRAMEWORK BASED ON BIO-INSPIRED OPTIMIZATION ALGORITHMS

GUILHERME FELIPPE PLICHOSKI

JOINVILLE, 2019

SANTA CATARINA STATE UNIVERSITY - UDESC
COLLEGE OF TECHNOLOGICAL SCIENCE - CCT
GRADUATE PROGRAM IN APPLIED COMPUTING - PPGCA

GUILHERME FELIPPE PLICHOSKI

A FACE RECOGNITION FRAMEWORK BASED ON BIO-INSPIRED OPTIMIZATION ALGORITHMS

JOINVILLE

2018

GUILHERME FELIPPE PLICHOSKI

A FACE RECOGNITION FRAMEWORK BASED ON BIO-INSPIRED OPTIMIZATION ALGORITHMS

Master thesis submitted to the Computer Science Department at the College of Technological Science of Santa Catarina State University in fulfillment of the partial requirement for the Master's degree in Applied Computing.

Advisor: Dr. Rafael Stubs Parpinelli
Co-advisor: Dr. Chidambaram Chidambaram

JOINVILLE

2018

Ficha catalográfica elaborada pelo programa de geração automática da Biblioteca Setorial do CCT/UDESC,

com os dados fornecidos pelo(a) autor(a)

Plichoski, Guilherme Felippe
A Face Recognition Framework based on Bio-Inspired Optimization Algorithms / Guilherme Felippe Plichoski. -- 2019.
95 p.

Orientador: Rafael Stubs Parpinelli
Coorientador: Chidambaram Chidambaram
Dissertação (mestrado) -- Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Computação Aplicada, Joinville, 2019.

1. Illumination variation. 2. Particle Swarm Optimization. 3. Differential Evolution. 4. Swarm Intelligence. 5. Evolutionary Algorithms. I. Parpinelli, Rafael Stubs. II. Chidambaram, Chidambaram. III. Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Computação Aplicada. IV. Titulo.

A FACE RECOGNITION FRAMEWORK BASED ON BIO-INSPIRED OPTIMIZATION ALGORITHMS

por

Guilherme Felippe Plichoski

Esta dissertação foi julgada adequada para obtenção do título de

Mestre em Computação Aplicada

Área de concentração em "Ciência da Computação",
e aprovada em sua forma final pelo

CURSO DE MESTRADO ACADÊMICO EM COMPUTAÇÃO APLICADA
DO CENTRO DE CIÊNCIAS TECNOLÓGICAS DA
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.

Banca Examinadora:

Prof. Dr. Rafael Stubs Parpinelli
CCT/UDESC (Orientador/Presidente)

Prof. Dr. Gilmário Barbosa
CCT/UDESC

Prof. Dr. Hugo Alberto Perlin
IFPR

Joinville,SC, 14 de dezembro de 2018.

I dedicate this to my mother, father, my life partner, Priscila, and my family in general; *amare et sapere vix deo conceditur*.

ACKNOWLEDGEMENTS

I would like to dedicate this work to my family, friends, colleagues and teachers who gave me the support when I needed and tuition to accomplish this master's thesis. Specially, my mother that helped me in every way she possibly could, taking from her and giving to me. Thanks for the advice when I needed and the financial support. The same goes to my father that was there when I needed. My friend Germano for always say to me the importance of studying. My life partner, Priscila, for being comprehensive and patient, supporting me as she could. Without you, I would not have the strength and discipline to achieve this. Dr. Chidambaram for teaching me, not only how to perform a scientific research, but also how behave professionally, and helping me open the pathways to achieve this master's degree since my graduation days. Dr. Rafael Parpinelli for being tough with me, this way, awakening all the potential in me. Finally, my grandpa Waldemiro for being my role model since I was a child, your impeccable conduct and your principles guided me to be the person I am today.

Furthermore, I would like to thanks for the financial support by the federal government agency CAPES and the Santa Catarina state government agency FAPESC from Brazil.

“All things by immortal power, near or far, hiddenly to each other linked are, that thou canst not stir a flower without troubling of a star.”

Francis Thompson

ABSTRACT

From current literature, it is well known that Face Recognition (FR) systems cannot perform well under uncontrolled conditions, thus there are no general and robust approaches with total immunity to all conditions. Hence, we present an adaptive FR framework with the aid of bio-inspired optimization algorithms. This approach implements several preprocessing and feature extraction techniques. The main feature of the present work stands on the use of the optimization algorithm which is responsible for choosing which strategies to use, as well as tuning their parameters. In this work, we employed three optimization algorithms for this task, namely Particle Swarm Optimization (PSO), Differential Evolution (DE) and the Self-adaptive Differential Evolution (jDE), aiming to address the illumination compensation problem. According to the proposed FR framework, the optimization algorithm can choose any set of the following techniques and tune their parameters: the Gamma Intensity Correction (GIC), the Wavelet-based Illumination Normalization (WBIN), the Gaussian Blur, the Laplacian for edge detection, the Discrete Wavelet Transform (DWT), the Discrete Cosine Transform (DCT), and the Local Binary Patterns (LBP). Experimental results are achieved using three well known databases i.e. The Yale Extended B, CMU-PIE and FERET, which confirmed that the proposed approach is suitable for FR.

Keywords: Illumination variation, Particle Swarm Optimization, Differential Evolution, Swarm Intelligence, Evolutionary Algorithms.

RESUMO

Da literatura atual, sabe-se que sistemas de reconhecimento facial não funcionam efetivamente sob condições não controladas, não havendo abordagens robustas com total imunidade a todas as condições. Com isto, apresentamos neste trabalho um *framework* de reconhecimento facial adaptativo que utiliza algoritmos de otimização bio-inspirados. A abordagem proposta implementa várias técnicas de pré-processamento e extração de características. O principal aspecto do presente trabalho é a utilização do algoritmo de otimização, que é responsável por escolher quais técnicas usar, bem como ajustar seus os parâmetros envolvidos. Neste trabalho, foram empregados três algoritmos de otimização, i.e. *Particle Swarm Optimization (PSO)*, *Differential Evolution (DE)* e uma versão auto-adaptativa do *Differential Evolution* (jDE), visando resolver o problema da variação de iluminação. De acordo com o *framework* proposto, o algoritmo de otimização pode escolher qualquer conjunto das seguintes técnicas e ajustar seus respectivos parâmetros: o *Gamma Intensity Correction (GIC)*, o *Wavelet-based Illumination Normalization (WBIN)*, o *Gaussian Blur*, o Laplaciano para detecção de bordas, o *Discrete Wavelet Transform (DWT)*, o *Discrete Cosine Transform (DCT)* e o *Local Binary Patterns (LBP)*. Os resultados foram obtidos utilizando três bases de imagens bem conhecidas i.e. Yale Extended B, CMU-PIE e FERET, e confirmaram que a abordagem proposta é adequada para reconhecimento facial.

Palavras-chave: Variação de iluminação, Particle Swarm Optimization, Differential Evolution, Inteligência de Enxames, Algoritmos Evolucionários.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ABC Artificial Bee Colony

ABPSO Accelerated Binary Particle Swarm Optimization

ACO Ant Colony Optimization

ADMM Alternating Direction Method of Multipliers

AGFN Gabor Feedforward Network

AMBPSO Adaptive Multi-Level Threshold Binary Particle Swarm Optimization

AR Aleix Martinez and Robert Benavente

ASE Academic Search Engine

BA Bees Algorithm

BFO Bacterial Foraging Optimization

BPSO Binary Particle Swarm Optimization

CAS-PEAL Academy of Sciences pose, expression, accessories, and lighting

CZT Chirp Z-Transform

CL Classification

CMU-PIE Carnegie Mellon University pose, illumination and expression

CNN Convolutional Neural Networks

CSA Cuckoo Search Algorithm

CT Contourlet Transform

DCT Discrete Cosine Transform

DD-DTCWT Double-Density Dual-Tree Complex Wavelet Transform

DE Differential Evolution

DFT Discrete Fourier Transform

DWT Discrete Wavelet Transform

EBPSO Exponential Binary Particle Swarm Optimization

EC Evolutionary Computation

ELM Extreme Learning Machine

ESRC Extended Sparse Representation Classifier

FEI Fédération Equestre Internationale

FERET Facial Recognition Technology

FFT Fast Fourier Transform

FMT Fourier-Mellin Transform

FR Face Recognition

FS Feature Selection

FSPSO Fast Static Particle Swarm Optimization

GA Genetic Algorithms

GAP Grayscale Arranging Pairs

GIC Gamma Intensity Correction

GRBRS Gamma Ray Burst Rhombus Star

GSA Gravitational Search Algorithm

GT Georgia Tech

GWT Gabor Wavelet Transform

H&L High-and Low-frequency Additive

HF Homomorphic filter

HH Diagonal component from 2D Discrete Wavelet Transform

HL Horizontal component from 2D Discrete Wavelet Transform

HMM Hidden Markov Model

HPC High Performance Computing

HT Hough Transform

ICA Intrinsic Discriminant Analysis

IC&CI Intelligent Control & Computational Intelligence Laboratory

IFD Indian Face Database

ISI-WoS Web of Knowledge

JAFFE Japanese Female Facial Expression

jDE Self-adaptive Differential Evolution

k-NN k-Nearest Neighbors

LBP Local Binary Patterns

LBPSO Logarithmic Binary Particle Swarm Optimization

LDA Linear Discriminant Analysis

LFW Labeled Faces in the Wild

LH Vertical component from 2D Discrete Wavelet Transform

LL Approximation component from 2D Discrete Wavelet Transform

LRSR Local Robust Sparse Representation

LTV Logarithmic Total Variation

MFSA Multiple Feature Subspaces Analysis

MIT-CBCL Massachusetts Institute of Technology - Center for Biological and Computational Learning

NMR Nuclear norm based Matrix Regression

OPSO Opposition Particle Swarm Optimization

ORL Olivetti Research Laboratory

OSPP One Sample Per Person

PCA Principal Component Analysis

PHPD Pointing Head and Pose Image Database

PO Parameter Optimization

PSO Particle Swarm Optimization

PUT Poznan University of Technology

R&L Reflectance and Illumination Additive

RBFNN Radial Basis Function Neural Network

RDWT Raster Scan Discrete Wavelet Transform

RQ Research Questions

RR-PSO Regressive Particle Swarm Optimization

SI Swarm Intelligence

SIFE Shift Invariance based Feature Extraction

SLPMM Supervised Locality Preserving Multimanifold

SLR Systematic Literature Review

SQI Self Quotient Image

STDCT Slope-form Triangular Discrete Cosine Transform

SVM Support Vector Machine

SWT Stationary Wavelet Transform

ThBPSO Threshold Binary Particle Swarm Optimization

TM Template Matching

TWEET Thresholded Wavelet Edge Enhancement Transform

UMIST University of Manchester Institute of Science and Technology

VITM Vikrant Institute of Technology & Management

WBIN Wavelet-based Illumination Normalization

WBPSO Weighted Binary Particle Swarm Optimizer

WDM Wavelet Denoising Model

WTFE Wavelet Transform Feature Extraction

LIST OF SYMBOLS

τ	Probability to adjust the control parameters of Self-adaptive Differential Evolution
γ	The parameter gamma in Gamma Intensity Correction technique
σ	Standard deviation of a given distribution
ϕ	Phi constant
bin	Binomial
c	A constant in Gamma Intensity Correction technique
c_1	Cognitive component in Particle Swarm Optimization
c_2	Social component in Particle Swarm Optimization
CR	Crossover probability in Differential Evolution algorithm
D_{image}	Database image in classification
$DE/a/b/c$	Representation of Differential Evolution configuration
$DistMeasure$	Distance Measure used in classification task
exp	Exponential
F	Mutation factor in Differential Evolution algorithm
F_{lower}	Mutation factor's lower bound
F_{upper}	Mutation factor's upper bound
$F(u,v)$	The output image after 2D Discrete Cosine Transform
$flag$	Boolean flag to represent if the technique will be employed or not
$G(x,y)$	The 2D Gaussian distribution
$I(x,y)$	Image in spatial domain
$L(x,y)$	The Laplacian of an image
L_∞	The L-Infinity norm
L_1	The Manhattan distance
L_2	The Euclidean distance
L_{2SQR}	The Euclidean distance squared
$level$	Number of levels of decomposition in 2D Discrete Wavelet Transform

k	Parameter k of k-Nearest Neighbors
MF	Most similar face image
$NMDF$	Maximum diversity obtained in the optimization process
$O(u, v)$	Gamma corrected image
p_{gbest}	Best position overall in Particle Swarm Optimization
$p - value$	Probability value
p_i	Particle's position in Particle Swarm Optimization
p_i^{best}	Particle's best position in Particle Swarm Optimization
$Population_{size}$	Population size
$Problem_{size}$	Problem size
$rand()$	Random number in the interval [0, 1]
S_{image}	Searching image in classification
S_r	The samples used in Wilcoxon signed-rank test
sdX	standard deviation in X of Gaussian Blur technique
sdY	standard deviation in Y of Gaussian Blur technique
SF	Scale Factor
t	Iterations representation in optimization algorithms
v_i	Particle's velocity in Particle Swarm Optimization
V_{max}	Maximum velocity in Particle Swarm Optimization
w	Inertia weight in Particle Swarm Optimization
WF	Wavelet Family

CONTENTS

1 INTRODUCTION

Face Recognition (FR) systems are widely used in different parts of the society which includes, for example, residences, public places, industries, commercial shops and offices. Generally, biometric-based human recognition systems are becoming popular making possible to surpass traditional security systems such as passwords (ISLAM et al., 2012; LIU et al., 2018a). Hence, it is possible to find security systems based on biometric identification such as gait, fingerprint, signature, voice, iris, face recognition, and others (NEBTI; BOUKERRAM, 2017). Among these, FR systems stands out due to the necessity of an identification system for real-world situations with a high flow of people. As an example, it is impractical to stop everyone at the entrance of an airport to record their fingerprint, but to place a camera recording the flow is enough for face capture. It is well known that FR systems perform well under controlled settings (OCHOA-VILLEGAS et al., 2015), however they are still facing great challenges such as variations of illumination, pose, expression and occlusion (LIU et al., 2018b; ZHANG et al., 2018).

Several algorithms have been proposed and a huge progress in a variety of applications and research fields was made. Recent methods that are based on Convolutional Neural Networks (CNN) achieved a great progress but their performance is still not enough for some real-world applications due to problems such as data bias and overfitting (WEN et al., 2018). Thus, in particular contexts, FR remains an open research topic.

Among different kind of approaches proposed during the last decades, some studies can be found in the literature which are focused on approaches that employ optimization techniques inspired by Nature, i.e. bio-inspired optimization techniques (BOWYER; CHANG; FLYNN, 2006; ABATE et al., 2007; ISLAM et al., 2012; SCHEENSTRA; RUIFROK; VELTKAMP, 2005; KONG et al., 2005; ZHAO et al., 2003; ALSALIBI et al., 2015). Moreover, an increasing number of bio-inspired FR systems had emerged due their intelligent problem-solving ability, scalability, flexibility, and adaptive nature. Also, biologically motivated algorithms are taking place due to their intrinsic parallelism and simplicity in computation (DAS; ABRAHAM; KONAR, 2008). Swarm Intelligence (SI) and Evolutionary Computation (EC) are the two major branches representing the biologically inspired algorithms for optimization. SI-based approaches are inspired by social and collective behavior of insects, such as ants, termites, bees, flock of birds and fish school. In this branch, we can cite the Particle Swarm Optimization (PSO) algorithm (EBERHART; SHI, 2011), a well-known algorithm among researchers. It is inspired by the coordinated movement of fish schools and bird flocks. Among many versions of PSO, its binary version has been widely used to find the most discriminative set of features in facial images improving FR systems (VORA et al., 2014; VARUN et al., 2015; VARADARAJAN et al., 2015). Another popular SI-based algorithm is the Ant Colony Optimization (ACO) (DORIGO; STUTZLE, 2003), which is inspired by the collective behavior of ants in finding the shortest path between the nest and the food source using a substance called pheromone. This algorithm has been used for feature selection and recognition purposes in 2D FR systems (KANAN; FAEZ; HOSSEINZADEH, 2007; VENKATESAN; MADANE, 2010; KAUR; PANCHAL; KUMAR, 2013). Other SI algorithms that have been used in FR systems are inspired by bacterial foraging (Bacterial Foraging Optimization (BFO)) (PASSINO, 2002), bees foraging (Bees Algorithm (BA), Artificial Bee Colony (ABC)) (PHAM et al., 2011; KARABOGA, 2005), brood parasitic behavior of some cuckoo species (Cuckoo Search Algorithm (CSA)) (YANG; DEB, 2009), and gravity law and mass interactions (Gravitational Search Algorithm (GSA)) (RASHEDI; NEZAMABADI-POUR; SARYAZDI, 2009). Many other SI algorithms can be found in Parpinelli and Lopes (2011). EC-based algorithms are inspired by the evolutionary theory proposed by Darwin. In this branch we may cite the Genetic Algorithms (GA) (HOLLAND, 1973), in which natural selection and genetic operators play the main role, and has been used for feature selection and classification in FR systems (FAN; VERMA, 2004; ZHENG; LAI; YUEN, 2005; LIU; WANG, 2006) and Differential Evolution (DE) algorithm (STORN; PRICE, 1997) which has been used as an optimization technique in FR systems (MALLIPEDDI; LEE, 2012; YOO; OH; PEDRYCZ, 2013).

In this work, a FR framework is proposed with the aid of bio-inspired optimization algorithms. The intent of the proposed methodology is to provide techniques for preprocessing and for feature extraction in which the optimization process, carried out by the bio-inspired algorithm, gives as output the best set of techniques for the task. Besides that, the parameters of chosen techniques are also optimized by the framework. Due to the non-ideal imaging environments frequently found in practical cases, uncontrolled illumination becomes a critical factor for real-world applications (ZHANG et al., 2018). Hence, the focus of this work is to achieve robustness for illumination variations. Based on the literature and on the illumination compensation properties, the following techniques were employed to deal with the illumination variation, which are Gamma Intensity Correction (GIC), Wavelet-based Illumination Normalization (WBIN), Gaussian Blur, Laplacian for edge detection, Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) and Local Binary Patterns (LBP). For the optimization process, we notice that bio-inspired 2D FR systems are trending towards the PSO algorithm which suggests a reliable choice. Also, experiments were carried out with one of the most used EC-based algorithm i.e. DE for comparison purposes.

The performance of the proposed FR system strongly depends on the chosen techniques and their parameter tunning, as well as the challenges encountered in the databases.

Three case studies were performed using three well known databases found in literature. In the first study, the Yale B database was employed, which contains several degrees of illumination and its cropped version of face images is important to investigate the illumination variation problem. Hence, its extended version (GEORGHIADES; BELHUMEUR; KRIEGMAN, 2001) is used for experimentation aiming to work on the illumination variations contained in this dataset. In this case study, the PSO and DE algorithms were employed and their performance was compared.

For the other two case studies the CMU-PIE and FERET databases were employed. The CMU-PIE database (SIM; BAKER; BSAT, 2002) is split into 5 different subsets with pose variation among them and several degrees of illumination within each subset. The objective with this configuration is to verify the system's ability to extract features from face images with different poses across illumination variation. Furthermore, it was used a single sample per person in the training set. In many situations, it is difficult to collect abundant data e.g. driving license or visas, this way, the One Sample Per Person (OSPP) problem represents a challenging real-world condition since only limited face image information is available (GU; HU; LI, 2018). Also, experimentation using all CMU-PIE subsets grouped together (C_{All}) were also performed to verify the robustness of the proposed methodology when intra-pose variation is considered. From FERET database (PHILLIPS et al., 1998), the b-series subset was used which contains only high changes in pose angles (+60° to –60°). It is worth to point out that the OSPP problem is considered for this subset as well.

In the second case study, different parameter's configurations for DE were employed aiming to reduce processing time. The results achieved are evaluated and compared with DE's initial configuration. Finally, in the third case study, a self-adaptive version of DE called jDE (BREST; ZUMER; MAUCEC, 2006) was employed, since it frees the user from the task of adjusting the control parameters. For this, the same experimental methodology using CMU-PIE and FERET was applied. All the results obtained are compared with related work found in the literature.

1.1 MOTIVATION

Some challenges arise in FR systems such as occlusion, scale, pose variants and uncontrolled illumination. Among these conditions that affect the FR process, illumination variation can be considered as one of the most critical (ZHANG et al., 2018). It affects the classification rate in a more significant manner than even the effect caused by having different subjects in a database (ADINI; MOSES; ULLMAN, 1997). According to the review performed in Ochoa-Villegas et al. (2015), it is possible to identify many illumination compensation techniques that perform well on images obtained under uncontrolled conditions, but there is no generic technique with total immunity to overall conditions. This context strongly implies that there is more space for improvement and investigation to develop new approaches.

1.2 OBJECTIVES

The main goal of this work is to develop a system that is generic enough to deal with the aforementioned challenges found in FR systems. To accomplish that, **(a)** the search mechanism intrinsic in the optimization algorithm used for training must be efficient to do such task, and **(b)** the implemented techniques must be robust enough to deal with these problems.

Following, some secondary objectives are listed, which will build the steps to accomplish the main objective:

- To implement a set of techniques that deal with the aforementioned problems.
- To develop the system employing bio-inspired optimization algorithms, such as DE and PSO.
- To analyze the system's performance using each of the optimization algorithms on datasets with uncontrolled illumination.

1.3 DOCUMENT STRUCTURE

The remaining chapters of this master's thesis are organized as follows: Chapter 2 presents the theoretical foundation required to comprehend this work, introducing the techniques related to FR, and the bio-inspired algorithms; Chapter 3 presents the SLR, concerning bio-inspired algorithms employed in FR systems; Chapter 4 provides a brief explanation about the related work; Chapter 5 depicts the proposed FR system; Chapter 6 provides the information about how the experiments are carried out, a discussion about the optimization algorithms performances, and a comparison between the results obtained; finally, Chapter 7 presents the conclusions, some future directions and the contributions made during the development of this master thesis.

2 BACKGROUND

In this chapter, we present the necessary theoretical foundation to understand this work. Also, the techniques employed to deal with the FR problems are described. Finally, we present a detailed explanation about the bio-inspired algorithms employed in this study, namely Differential Evolution (DE), Self-adaptive Differential Evolution (jDE) and Particle Swarm Optimization (PSO).

2.1 FACE RECOGNITION

Traditional security systems based on encryption and passwords have proven vulnerable and easily breakable. Hence, the biometric technology became crucial for many domains (NEBTI; BOUKERRAM, 2017). Among them, 2D-FR systems do not require any user interaction which becomes an advantage compared with other biometric approaches (KIM; OH; KIM, 2016). However, some issues arise on image acquisition such as occlusion, different pose and expressions and illumination variation. Following, we present the techniques employed in this work to deal with the illumination variation problem.

2.1.1 Preprocessing Techniques

In this section, the chosen preprocessing techniques are exposed, as well as their parameters.

2.1.1.1 Wavelet-based Illumination Normalization (WBIN)

Wavelet decomposition has been successfully applied in the illumination normalization problem due its ability to provide a multi-resolution analysis of the image and the capability of decomposing the image into sub-bands in both time and frequency domains. The technique for illumination normalization proposed by Du and Ward (2005) was implemented, namely Wavelet-based Illumination Normalization (WBIN). In the 2D-DWT, the image is represented in terms of translations and dilations of a scaling function and a wavelet function using a 2D filter bank consisting of low-pass and high-pass filters (DU; WARD, 2005). The different wavelet families (WF) make different trade-offs between how compactly the basis functions are localized in space and how smooth they are, which are biorthogonal (bior), reversed biorthogonal (rbio), coiflets (coif), daubechies (db), haar and symlets (sym). The biorthogonal wavelet exhibits the properties of linearity which is advantageous for image and signal reconstruction, while the reverse biorthogonal wavelet is obtained by biorthogonal wavelet pairs. The daubechies family is an orthogonal wavelet and its key characteristic is the availability of maximum number of vanishing moments for some predefined support length. The symlets and coiflets are daubechies with stronger symmetry. The haar wavelet was the first mother wavelet proposed and is known for having the shortest length of support among all orthogonal wavelets (DOGRA, 2017). Each family contains several variations according to its smoothness (GRAPS, 1995). High pass filtering produces detail information (such as edges) and low pass filtering with scaling produces coarse approximations. The 2D-DWT produces four components, which are approximation (LL), horizontal (HL), vertical (LH), and diagonal (HH), as shown in Figure 1.

Figure 1 – DWT decomposition levels. Left: original image. Middle: Decomposed components. Right: Components specification.

Source: Own authorship

For the illumination normalization procedure using WBIN, histogram equalization is normally 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 Scaling Factor (SF). Then, the enhanced image is reconstructed using the inverse wavelet transform. The parameters considered to this technique are: the WF that can assume 105 possible configurations and the SF that can assume any value greater than 0. A list with all possible values for the WF parameter is shown in Annex A.

2.1.1.2 Gamma Intensity Correction (GIC)

The Gamma Intensity Correction (GIC) technique was implemented since it is useful to correct image's luminance. According to Shan et al. (2003), gamma correction, also known as power law transformation, can control overall brightness of an image by changing the parameter γ. It's formula is shown in Equation 2.1.

$$O(u,v) = c * I(x,y)^{\gamma} \tag{2.1}$$

where $O(u, v)$ is the gamma corrected image, c is a constant, $I(x, y)$ is the input image and γ is the correction factor (MANIKANTAN et al., 2012). The only parameter to adjust within this technique is the parameter γ. If it assumes values lower than 1 it will shift the image towards the darker end of the spectrum, while γ values greater than 1 will make the image brighter.

2.1.1.3 Gaussian Blur

Gaussian Blur was implemented in the FR framework. This technique uses a 2D convolution operator with a Gaussian distribution to blur and remove detail and noise from the image (GONZALEZ; WOODS, 2008). The 2D Gaussian distribution is shown in Equation 2.2.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.2}$$

where σ is the standard deviation of the distribution. The parameters to this technique are: The Gaussian kernel size, which must assume positive and odd values, and Gaussian kernel standard deviation in X (sdX) and Y (sdY) directions (ITSEEZ, 2018).

2.1.1.4 Laplacian for Edge Detection

The edge detection includes several mathematical methods which identify the image points that sharply changes in brightness, thus representing the shape of face, eyes, nose and other discriminant features. The Laplacian operator is one of the methods that is commonly used to identify the edges. In Kushwah et al. (2017), a review of edge detection techniques was performed which featured Laplacian with better results for visual perception and edge quantity. It is worth to point out that this technique usually has better performance along with a technique that removes noise such as Gaussian Blur (Section 2.1.1.3).

The Laplacian $L(x, y)$ of an image with pixel intensity values $I(x, y)$ is given by Equation 2.3

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{2.3}$$

And can be calculated using the following convolution filter,

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

The parameters considered to this technique are: the aperture size that is used to compute the second-derivative filters, which must be positive and odd, a SF for the computed Laplacian values and a delta value that is added to the results prior to storing them in resulting image (ITSEEZ, 2018).

2.1.2 Feature Extraction Techniques

It was possible to infer from literature that a large number of works employed the Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) techniques for feature extraction, suggesting a reliable choice for its usage in FR systems. In addition, an important characteristic to describe surfaces is texture. That said, the texture descriptor Local Binary Patterns (LBP) is a highly used technique for feature extraction. In the following subsections, we explain about each feature extraction technique, as well as its parameters.

2.1.2.1 2D Discrete Wavelet Transform (DWT)

As mentioned in Section 2.1.1.1, the 2D-DWT produces four sub-bands (Figure 1) containing the approximation component (LL) and three detail components (HL, LH and HH). According to Manikantan et al. (2012), the information in the approximation component has the most discriminant features. Hence, only this sub-band was used. This procedure may have n levels of decomposition. The parameters considered to this technique are: the WF that can assume 105 possible configurations as mentioned in Section 2.1.1.1 and the number of levels of decomposition ($level$). A list with all possible values for the WF parameter is shown in Annex A.

2.1.2.2 2D Discrete Cosine Transform (DCT)

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 tends to be concentrated in low frequency components or at corner of the spectrum (RAO; RAO, 2016). Equation 2.4 shows the DCT procedure for an image pixel value of $I(x, y)$ of dimension NxM in the spatial domain. The output image in the frequency domain is represented by $F(u, v)$.

$$F(u,v) = \alpha_u \alpha_v I(x,y) \cos\left(\frac{\pi(2x+1)u}{M}\right) \cos\left(\frac{\pi(2y+1)v}{N}\right) \tag{2.4}$$

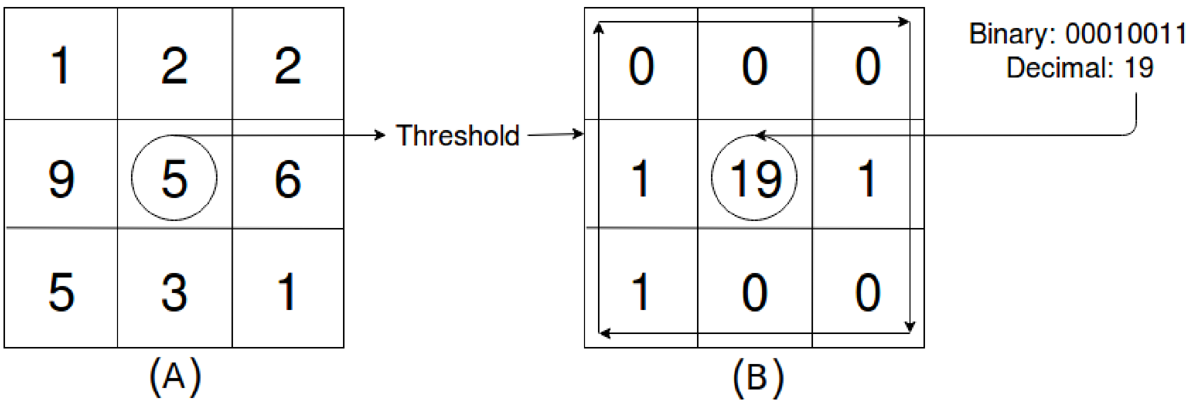
since,

$$\alpha_v \begin{cases} \frac{1}{\sqrt{N}}, u = 0 \\ \sqrt{\frac{2}{N}}, 1 \leq u \leq N-1 \end{cases} \qquad \alpha_u \begin{cases} \frac{1}{\sqrt{M}}, u = 0 \\ \sqrt{\frac{2}{M}}, 1 \leq u \leq M-1 \end{cases}$$

2.1.2.3 Local Binary Patterns (LBP)

When choosing a texture descriptor, two competing goals should be considered that are low computational complexity and capturing the most representative texture information. According to Liu et al. (2017), the advantages of this technique are its easy of implementation, invariance to monotonic illumination changes and low computational complexity. Its canonical version is used in this work in which the features are extracted directly from the input image. Figure 2 illustrates LBP's procedure with a kernel size (*ksize*) of 3x3.

Figure 2 – LBP texture descriptor procedure.



Source: Own authorship

The neighborhood structure is a set of pixels taken from a square neighborhood, which are compared against the value of the central pixel (A) resulting in a 8 bits binary vector which is converted to decimal. Then, the resultant value is used to represent the pixel (B).

2.1.3 Classification

For the classification task, an image to be classified is compared to each and every image in the database in terms of similarity returning the most similar one, denoted by MS. Then, a match occurs if both images are from the same person. This classification is carried out by the k-Nearest Neighbors (k-NN) classifier, where $k = 1$ (COVER; HART, 1967). Equation 2.5 represents this comparison, where S_{image} represents the searching image, D_{image}, the database image and $DistMeasure$, the distance measure. In this work, the distance measure is considered as a parameter, thus, we included it in our optimization task.

$$MS = \underset{1 \leq D_{image} \leq N}{\operatorname{argmin}} \left(DistMeasure(S_{image}, D_{image}) \right) \tag{2.5}$$

The first distance measurement implemented is the Manhattan distance (L_1), that is the distance between two points alongside the axis at right angles, as shown in Equation 2.6. The known Euclidean distance (L_2) was also employed according to Equation 2.7. Third, L_{2SQR} was added (Equation 2.8). Finally, L_∞ was implemented according to Equation 2.9.

$$DistMeasure(x, y) = \sum_{i=1}^{N} |x_i - y_i| \tag{2.6}$$

$$DistMeasure(x, y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{2.7}$$

$$DistMeasure(x, y) = \sum_{i=1}^{N} (x_i - y_i)^2 \tag{2.8}$$

$$DistMeasure(x, y) = \max |x_i - y_i| \tag{2.9}$$

2.2 BIO-INSPIRED ALGORITHMS

Bio-inspired algorithms can be classified into five different groups based on their source of inspiration, which are physico-chemical, Neural Networks, Immunological Systems, Swarm Intelligence (SI) and Evolutionary Computation (EC) (Fister JR. et al., 2013; ANDRÉ; PARPINELLI, 2014). Algorithms inspired by physiochemical systems were designed to mimic the behavior and characteristics of certain laws of physics or chemistry, including gravity, electric charges, and pluvial systems. In this group we can cite Big-Bang Big-Crunch Optimization (EROL; EKSIN, 2006), Black Hole Optimization (HATAMLOU, 2013), Simulated Annealing (KIRKPATRICK; GELATT; VECCHI, 1983), among others (Fister JR. et al., 2013). Algorithms based on biological systems have its source of inspiration originated from Biology. In this category we have Artificial Neural Networks (MCCULLOCH; PITTS, 1943), Artificial Immune Systems (DASGUPTA; YU; NINO, 2011), EC (JONG, 2006) and SI (PARPINELLI; LOPES, 2011), being the last two groups the focus of this work due to their huge applicability to FR systems.

EC is based on the natural selection theory proposed by Darwin. Individuals of a population compete with each other in which the more adapted ones have higher reproductive chances to survive. Genetic Algorithms (GA) (HOLLAND, 1973) and Differential Evolution (DE) (STORN; PRICE, 1997) are some representatives in this group. SI algorithms are inspired by the social and collective behavior of insects, such as ants, termites, bees, flock of birds and fish school. The collective and self-organized behavior that appears from local interactions is the intelligence found in those systems, which is called emergent behavior. Ant Colony Optimization (ACO) (DORIGO; STUTZLE, 2003), Artificial Bee Colony (ABC) (KARABOGA, 2005), Bacterial Foraging Optimization (BFO) (PASSINO, 2002), Bees Algorithm (BA) (PHAM et al., 2011), Cuckoo Search Algorithm (CSA) (YANG; DEB, 2009), Gravitational Search Algorithm (GSA) (RASHEDI; NEZAMABADI-POUR; SARYAZDI, 2009), and Particle Swarm Optimization (PSO) (EBERHART; SHI, 2011) are some representatives in this group.

A common feature present in EC and SI algorithms is that they are population-based approaches. In a population-based algorithm the optimization process takes place in a set of candidate solutions at each iteration. This set of solutions can be called population, swarm, school, or hive, depending on the biological inspiration employed, and each corresponding candidate solution can be an individual, a particle, a bee, or an ant. Although all these algorithms are very similar, they differ in the mechanism or criterion for selecting solutions and in the way they modify (or build) a candidate solution by applying intensification and diversification procedures. Hence, a general pseudo-code for population-based algorithms is shown in Algorithm 1 (BENITEZ; PARPINELLI; LOPES, 2012; FISTER; BREST; MLAKAR, 2016). In line 1 of the algorithm, a population of candidate solutions is randomly initialized, and in line 2 this initial population is evaluated according to the problem being solved. The main loop (between lines 3 to 9) represents the generations (or iterations) of the algorithm, and line 4 defines the mechanism for selecting solutions for next step (i.e., survival of the fittest in GA, define which solutions will guide the search in PSO, or define if the trial individuals replace the targets in DE). Two important characteristics of population-based algorithms, and also for meta-heuristics in general, are the intensification and diversification procedures used to produce new candidate solutions. In line 5 of the algorithm, intensification intends to search locally the search space (i.e., a crossover procedure in GA and DE, or the cognitive component in PSO), whereas diversification leads the algorithm to explore globally the search space (i.e., a mutation procedure in GA and DE, or the social component in PSO). In line 6 the new pool of candidate solutions is evaluated. In line 7 a replacement criterion can be applied and the main loop restarts until a termination condition is met. It is also common to these algorithms identify the current best solution at each iteration as pointed in line 8. Finally, line 10 outputs the best solution found in the optimization process.

Algorithm 1 General pseudo-code for population-based algorithms (BENITEZ; PARPINELLI; LOPES, 2012; FISTER; BREST; MLAKAR, 2016)

1: $Initialize$ the population with random candidate solutions
2: $Evaluate$ each candidate solution
3: **while** *stop condition is not reached* **do**
4:     Perform competitive $selection$
5:     Apply $intensification$ and $diversification$ procedures
6:     $Evaluate$ the new pool of candidate solutions
7:     Apply $replacement\ criterion$ to form the new population
8:     $Find$ current best solution
9: **end while**
10: **return** $Overall\ best\ solution;$

In this study, the DE and PSO algorithms are explored as part of the proposed FR framework. Despite their canonical versions, a self-adaptive version of DE called jDE was also employed. Next, the algorithms are detailed as well as their pseudo-codes.

2.2.1 Differential Evolution

Differential Evolution (DE) was proposed by Storm (1997) in 1997. It is a global optimization algorithm that encodes candidate solution vectors (or individuals) in real-valued numbers.

Initially, the individuals are generated randomly in the search space. New individuals are created adding the weighted difference between two individuals to a third one, namely target. This routine is called mutation. Then, the target individual is combined with a randomly preselected individual resulting the trial individual and representing crossover procedure. The decision if the trial individual replaces the target in the next generation is based on an objective function and its constraints (fitness). If trial vector is worse than the target's fitness value, the trial is discarded. Otherwise, the trial vector replaces the target. This operation is called greedy selection. This process repeats until a stop criterion is reached. The input parameters that enable to control the search are the population size, the crossover probability (CR) and the mutation factor (F).

Algorithm 2 shows DE pseudo-code considering a minimization problem (BROWNLEE, 2011). Algorithm 3 shows the procedure to generate a trial solution. In Algorithm 2, the population is initialized randomly and the individuals are evaluated based on fitness (lines 1-3), then this population is submitted to continuous optimization until a stop condition is met (lines 4-17). In this loop, a new population is generated based on mutation and crossover routines (line 7), which brings us to Algorithm 3. In this, individuals are randomly selected from the population to create the target (lines 1-9), and the crossover procedure is performed (lines 10-18). Following, lines 8-12 perform greedy selection between individuals of the two populations generating a new one for the next generation. Finally, when loop is over, the algorithm returns the overall best individual (line 18). DE has a nomenclature that describes the adopted configuration. This takes the form of $DE/a/b/c$, where a represents the solution to be perturbed, such as random or 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 (BROWNLEE, 2011). The most popular configuration is demonstrated in the referenced pseudo-codes i.e. $DE/rand/1/bin$.

2.2.2 Self-adaptive Differential Evolution (jDE)

The canonical DE keeps the control parameters fixed during the whole optimization process and are defined by the user. On the other hand, the self-adaptive version jDE (BREST; ZUMER; MAUCEC, 2006) 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 (BREST; ZUMER; MAUCEC, 2006).

The parameters F_i^{G+1} and CR_i^{G+1} are calculated as shown in Equation 2.10 and Equation 2.11 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.

$$F_i^{G+1} = \begin{cases} F_{lower} \ + \ rand_1 \ \cdot \ F_{upper} & \text{if } rand_2 < \tau_1, \\ F_i^G & \text{otherwise.} \end{cases} \tag{2.10}$$

Algorithm 2 Pseudo-code for Differential Evolution algorithm (BROWNLEE, 2011)

Input: $Population_{size}, Problem_{size}, F, CR$
Output: s_{best}
$Population \leftarrow InitializePopulation(Population_{size},$
$Problem_{size})$
$EvaluatePopulation(Population)$
$s_{best} \leftarrow GetBestSolution(Population)$
while ***stop condition is not reached*** **do**
  $NewPopulation \leftarrow \emptyset$
  for $p_i \in Population$ **do**
    $s_i \leftarrow NewSample(p_i, Population, Problem_{size},$
    $F, CR)$
    if $Cost(s_i) \leq Cost(p_i)$ **then**
      $NewPopulation \leftarrow s_i$
    else
      $NewPopulation \leftarrow p_i$
    end if
  end for
  $Population \leftarrow NewPopulation$
  $EvaluatePopulation(Population)$
  $s_{best} \leftarrow GetBestSolution(Population)$
end while
return s_{best}

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 $\vec{p}_i^{\,G+1}$.

$$CR_i^{G+1} = \begin{cases} rand_3 & \text{if } rand_4 < \tau_2, \\ CR_i^G & \text{otherwise.} \end{cases} \tag{2.11}$$

2.2.3 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) was proposed by Eberhart and Shi (2011) in 1995. This algorithm is population-based and simulates the coordinated behavior of a flock of birds or a fish school. Algorithm 4 (BROWNLEE, 2011) shows the pseudo-code of PSO to minimize a cost function. The algorithm initiates with a random population of particles (potential solutions). Each particle has its own velocity (lines 1-10), which is responsible for moving the particles around the search space. At each iteration the velocities are updated based on the best position of the particle so far (cognitive component) and the best position found globally (social component) (line 13). The update is done through Equation 2.12.

Algorithm 3 Pseudo-code for $NewSample$ function (BROWNLEE, 2011)

Input: $p_o, Population, Problem_{size}, F, CR$
Output: s

1: **repeat**
2: $\quad p_1 \leftarrow RandomMember(Population)$
3: **until** $p_1 \neq p_o$
4: **repeat**
5: $\quad p_2 \leftarrow RandomMember(Population)$
6: **until** $p_2 \neq p_o \vee p_2 \neq p_1$
7: **repeat**
8: $\quad p_3 \leftarrow RandomMember(Population)$
9: **until** $p_3 \neq p_o \vee p_3 \neq p_2 \vee p_3 \neq p_1$
10: $CutPoint \leftarrow RandomPosition(Problem_{size})$
11: $s \leftarrow 0$
12: **for** i **to** $Problem_{size}$ **do**
13: $\quad$ **if** $i \equiv CutPoint \wedge Rand() < CR$ **then**
14: $\quad\quad s_i \leftarrow p_{3_i} + F \times (p_{1_i} - p_{2_i})$
15: $\quad$ **else**
16: $\quad\quad s_i \leftarrow p_{o_i}$
17: $\quad$ **end if**
18: **end for**
19: **return** s

$$v_i(t+1) = w \times v_i(t) + (c_1 \times rand() \times (p_i^{best} - p_i(t))) + (c_2 \times rand() \times (p_{gbest} - p_i(t))) \tag{2.12}$$

where $v_i(t+1)$ is the new velocity for the i^{th} particle, $v_i(t)$ represents the current particle's velocity, w is an inertia or momentum coefficient used to limit the change in velocity, c_1 and c_2 are acceleration constants, $rand()$ is a random number in the interval [0, 1], p_i^{best} is the particle best position, p_{gbest} is the best position know so far and $p_i(t)$ is the current particle's position. The acceleration constant c_1 regulates the step size of the particle towards the cognitive component, while c_2 regulates the step size towards the social component. As pointed in Zhang, Yu and Hu (2003) velocities should be clamped by a maximum velocity V_{max} specified by the user to control intensification and diversification. A larger V_{max} promotes diversification, while a smaller V_{max} promotes intensification in the search. The velocities are usually clamped as $v_i(t+1) \in (-V_{max}, V_{max})$.

The next step is to update particle's position (line 14), which is done through Equation 2.13.

$$p_i(t+1) = p_i(t) + v_i(t) \tag{2.13}$$

Where, $p_i(t+1)$ is the new particle's position, $p_i(t)$ is the current particle's position and $v_i(t)$ is the particle's velocity. If particle's position exceeds search space limits, the particle should be attached to the domain's boundaries. Now the particles are evaluated in terms of fitness, and then the process repeats until a stop criterion is reached.

Algorithm 4 Pseudo-code for Particle Swarm Optimization algorithm (BROWNLEE, 2011)

Input: $Problem_{size}, Population_{size}, w, c_1, c_2$
Output: p_{g_best}

$Population \leftarrow \emptyset$
$p_{g\_best} \leftarrow \emptyset$
for $i = 1$ **to** $Population_{size}$ **do**
  $p_{velocity} \leftarrow RandomVelocity()$
  $p_{position} \leftarrow RandomPosition(Population_{size})$
  $p_{p\_best} \leftarrow p_{position}$
  if $Cost(p_{p\_best}) \leq Cost(p_{g\_best})$ **then**
    $p_{g\_best} \leftarrow p_{p\_best}$
  end if
end for
while *stop condition is not reached* **do**
  for $p \in Population$ **do**
    $p_{velocity} \leftarrow UpdateVelocity(p_{velocity}, p_{g\_best}, p_{p\_best}, w, c_1, c_2)$
    $p_{position} \leftarrow UpdatePosition(p_{position}, p_{velocity})$
    if $Cost(p_{position}) \leq Cost(p_{p\_best})$ **then**
      $p_{p\_best} \leftarrow p_{position}$
      if $Cost(p_{p\_best}) \leq Cost(p_{g\_best})$ **then**
        $p_{g\_best} \leftarrow p_{p\_best}$
      end if
    end if
  end for
end while
return p_{g_best}

3 SYSTEMATIC LITERATURE REVIEW

According to Detroz, Hinz and Hounsell (2015), the traditional bibliographies are oriented by the researchers experience, which might lead to impartial results. Unlike, a Systematic Literature Review (SLR) aims to present a fair evaluation of a research topic, identifying and evaluating in a reliable and impartial manner all relevant researches using a trustworthy, rigorous, and auditable methodology (KITCHENHAM; CHARTERS, 2007). Also, it allows to summarize the benefits and limitations of a specific method (KITCHENHAM, 2004). Based on this, we decided to perform a SLR of 2D FR systems using biologically inspired approaches (considering both SI and EC algorithms) (PLICHOSKI; PARPINELLI; CHIDAMBARAM, 2018c). Due the lack of knowledge at that moment about what and how authors were employing bio-inspired algorithms in the FR process, we found necessary to perform this study. Hence, we describe the research method employed next.

3.1 RESEARCH METHOD

In Petersen et al. (2008), the SLR methodology is presented. The first step is to define the Research Questions (RQ) in order to identify and evaluate all available relevant works, which are presented as follows:

- RQ$_1$: What bio-inspired optimization algorithms are being applied to 2D FR systems?
- RQ$_2$: How bio-inspired optimization algorithms are being applied in 2D FR systems?

With the scope defined, the SLR is effectively done through the following steps: planning and conducting.

3.1.1 Planning the Review

After defining the RQs, the search for relevant works must be done. This work performs an automated search, in which it is defined a boolean search string with keywords used as input in Academic Search Engines (ASEs). The following words were employed: faces, face, facial, recognition, detection, bioinspired, bioinspiration, bio-inspired, bio-inspiration, bio inspired, bio inspiration, evolutionary, and swarm, as well its synonyms and variations. The ASEs were chosen according to CAPES (2013), which represent the most relevant for Computer Science. The ASEs selected were Web of Knowledge (ISI-WoS), SCOPUS and IEEEXplore. The following boolean composition was used to perform the search: (FACES OR FACE OR FACIAL) AND (RECOGNITION OR DETECTION) AND (BIO-INSPIR* OR “BIO INSPIR*” OR BIOINSPIR* OR EVOLUTIONARY OR SWARM). However, as each ASE has its own search mechanism, this boolean query suffered some changes preserving its semantic meaning. The research was performed on March 13th of 2018.

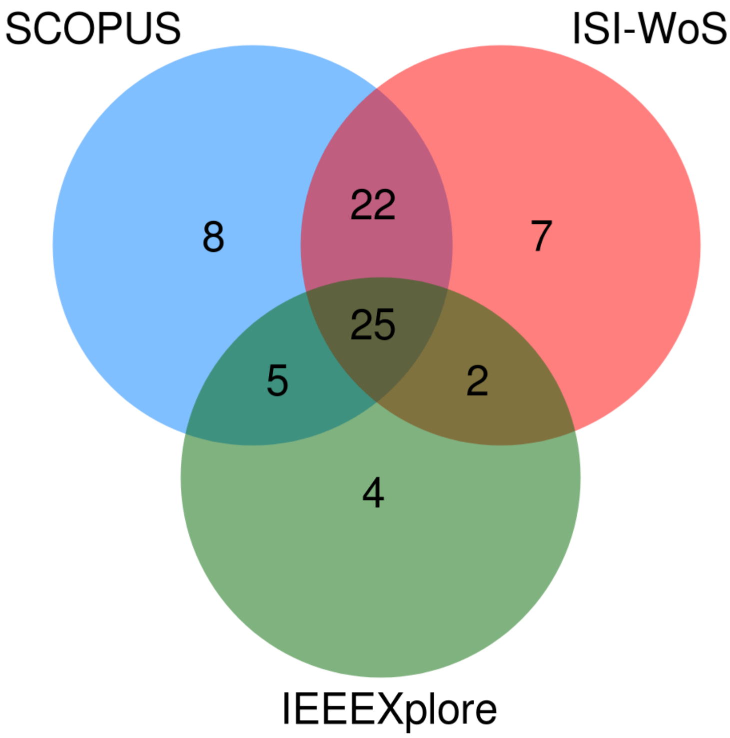
3.1.2 Conducting the Review

After defining the search engines and the boolean query, closure requirements are necessary, such as objective, inclusion, and exclusion criteria, which are presented as follows:

- Objective Criteria
  - Year range : 2012 - 2018
  - Document type : Articles
  - Language : English
  - Availability : Any
  - Completeness : Full or short
  - Duplicates : Yes
- Inclusion Criteria
  - To include works if, and only if it uses any swarm intelligence and/or evolutionary algorithms applied to FR systems.
- Exclusion Criteria
  - Remove works based on video recognition.
  - Remove works which aims to recognize only pose or emotions.
  - Remove works which use thermal face images.
  - Remove works which fuses features from other biometric techniques with face features.
  - Remove works which uses 3D face features.
  - Remove duplicated works.

Based on the above rules, it was established that if a work fits into any exclusion criteria or does not fit into any objective criteria, then it should be excluded. The evaluation of each paper returned by the boolean query was done in this order: objective, exclusion, and inclusion criteria. Figure 3 shows the relevant works found for each ASE.

Figure 3 – Relevant works found for each ASE.



Source: Own authorship

After removing the duplicated works, about seventy three papers were kept for analysis. Following, the relevant information identified through this research and a concise discussion are presented.

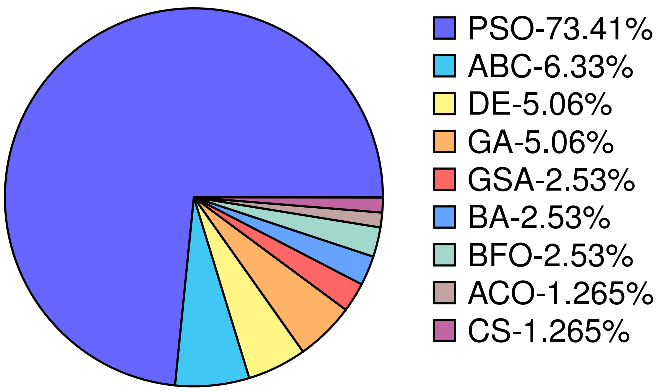
3.2 SUMMARY AND DISCUSSIONS

Most of the bio-inspired algorithms in 2D FR systems have been applied to Feature Selection (FS), Parameter Optimization (PO), Classification (CL) and Template Matching (TM). According to Rao and Rao (2016), the feature selection technique in a FR system consists in extracting the best features subset from the original images dataset, aiming to improve recognition rate. Many authors employed bio-inspired algorithms to optimize the intrinsic parameters in their proposed methodologies, such as selecting the parameters G and C of Support Vector Machine (SVM) classification (VALUVANATHORN; NITSUWAT; HUANG, 2012), searching the optimum Hidden

Markov Model (HMM) states and parameters (FARAG; ELGHAZALY; HEFNY, 2016), or adjusting the parameters of a homomorphic filter (PLICHOSKI; PARPINELLI; CHIDAMBARAM, 2017). In addition to these, there are other bio-inspired approaches that are employed for template matching, which consist in finding areas of an image that better match to a template image (CHIDAMBARAM et al., 2014). As preprocessing step, template matching might deal with problems such as scale and rotation variations. Finally, works were found for classification purposes, such as Nebti and Boukerram (2017) that employed a bio-inspired approach to classify a decision tree recursively until obtaining only one class representing the input face image, thus addressing illumination, pose and facial expression variations. Seventy three scientific articles were analyzed focusing on the problem approached, which bio-inspired algorithm are used, how candidate solutions are represented, how fitness function is modeled, and what databases are employed.

Figure 4 presents the distribution of bio-inspired algorithms among the analyzed works. A huge gap between the use of the PSO algorithm (73.41%) in comparison to other algorithms (26.59%) can be noticed. Some reasons for the popularity of PSO algorithm in 2D FR systems might be because of its good performance, its implementation simplicity, and the use of few parameters to be tuned. Also, as pointed out by Alsalibi et al. (2015), PSO algorithm requires less training time, has good scalability and high convergence rate. According to an extensive review done in Zhang, Wang and Ji (2015), the number of publications related to PSO is the highest among others reaching around 1,000 per year. However, the few works developed using other bio-inspired algorithms provides room for exploiting their features.

Figure 4 – Bio-inspired algorithms distribution in relation to the analyzed works

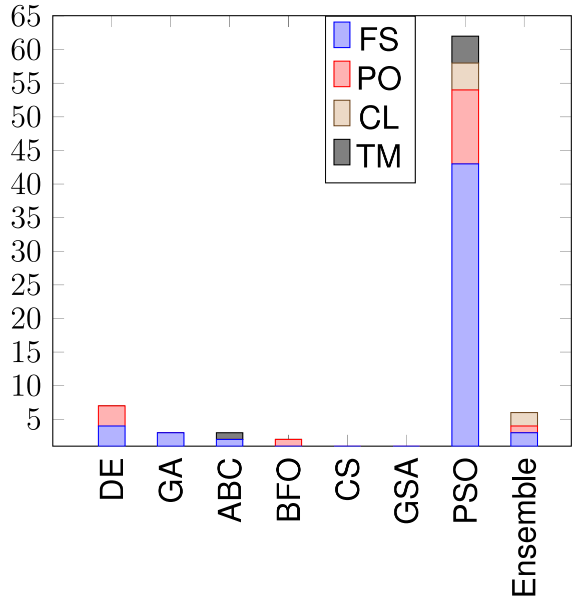


Source: Own authorship

Among the four applications employed by bio-inspired algorithms in 2D FR systems, the FS reached the highest rate of 76.315%, followed by PO with 19.74%.

In our review, we found only two works for classification and one work for template matching. Figure 5 shows the distribution of applications per algorithm in which the dominance of FS applications can be clearly observed. The high usage of FS can be justified since its use as a dimensional reduction approach becomes attractive and sometimes required. Additionally most of the works that deals with images and videos requires FS as an important step and it is an important component of many pattern recognition tasks with very high-dimensional data (GUI et al., 2017).

Figure 5 – Distribution of applications per algorithm



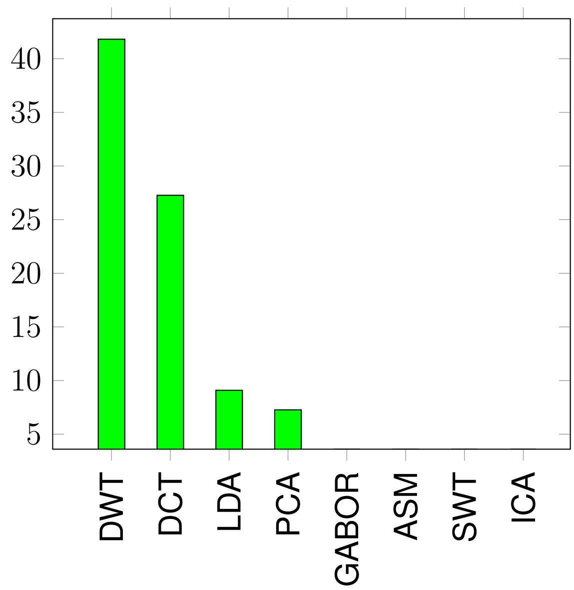
Source: Own authorship

For every work applied for FS, the candidate solution was represented as a vector containing the selected features. Also, in most works the search was driven aiming to maximize the distance between classes. The most used feature extraction techniques are presented in Figure 6. The DWT technique is the most employed (41.82%) mainly because of its good performance. Briefly, the DWT has the ability to provide a multi-resolution analysis of the image and the capability of decomposing the image into sub-bands in both time and frequency domains producing detailed information and coarse grained approximations (DU; WARD, 2005). Also, a number of works employing the DCT technique is expressive (27.27%). According to Rao and Rao (2016), the DCT is a popular transformation technique based on the energy compaction property, wherein the information or features tend to be concentrated in low frequency components or at the corner of the spectrum. Regarding the dimensionality reduction techniques, the LDA and PCA are popular in 2D FR systems with 9.09% and 7.27%, participation respectively. The other techniques represent in average about 3.6% each.

For the PO task, the candidate solution is represented as the parameters itself and in 56.25% of the works, the fitness function is constructed based on accuracy (recognition rate), followed by distance between classes maximization (18.75%). The most used techniques for tuning were the SVM (26.66%), the RBFNN (20%) and the LBP (13.33%).

For the CL task, distinct solution representations are employed. For example, one work used the image pixels as solution representation (KAUR; PANCHAL; KUMAR, 2013) and another used the database classes (NEBTI; BOUKERRAM, 2017). For the TM task, the individual is represented as a four-dimensional vector containing horizontal and vertical coordinates, scale factor and rotation angle. All works used the Euclidean distance measure to drive the search and to calculate distance between images (CHIDAMBARAM et al., 2014).

Figure 6 – Most used feature extraction techniques

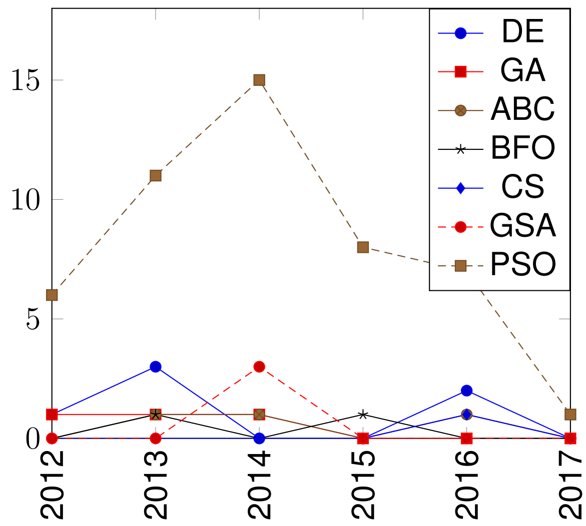


Source: Own authorship

Figure 7 shows the distribution of most used bio-inspired algorithms per year. The number of works employing PSO far exceeds all other algorithms. However, ABC and the set DE and GA occupy the second and third rankings respectively.

When analyzing the works presented in this review, we realize that the experimental methodology performed by each author usually differs, which sometimes makes it unreliable, difficult to compare with others and to replicate results. As an example, we can mention the use of the ORL database which contains 400 images of 40 individuals with 10 images of each one varying pose. Khadhraoui et al. (2016) used the first 5 images per person of all subjects for training and the remaining 5 images per person of all

Figure 7 – Number of bio-inspired algorithms employed per year



Source: Own autorship

subjects for testing. On the other hand, from the same database Fernández-Martínez and Cernea (2015) used 200 images for training, 160 images for testing and 40 images to perform blind validation. Despite that, we investigated the most used benchmark databases and are presented in Table 1. The complexity of the databases is defined by its size and categories such as pose, illumination, expressions and occlusion. The most used database is the FERET (representing 20.21% of the works) which contains 14,126 images with 512x768 pixels and 1,199 individuals varying pose, illumination and expressions (ALSALIBI et al., 2015). The next position is occupied by the ORL database with 17.55% and Yale B with 13.83% of the works. Yale B database is a well known database because of its several degrees of illumination and its cropped version of face images which is important to investigate the illumination variation problem. The remaining databases information are shown in references described in Table 1.

Nowadays, in different parts of the society, we can find a pool of 2D FR systems that perform well under controlled settings. However, real-time situations require systems that should deal with uncontrolled settings and thus creating an open research gap. The major challenges found in the 2D FR systems refers to the high-dimensionality of the dataset, a wide range of image variations and the useless information in the images leading to misclassification. Such complexity makes necessary the use of optimization methods as bio-inspired algorithms to construct robust FR systems. Hence, the researchers from the whole world have been proposed many 2D FR approaches using bio-inspired algorithms. Hence, in this context, to map these approaches we performed the present SLR. Based on this study and intrinsic illumination compensation

Table 1 – Most popular databases.

Database	Subjects	Total Images	Variations	Reference	Occurrence (%)
Facial Recognition Technology (FERET)	1,199	14,126	p,i,e	(PHILLIPS et al., 2000)	20.21
Olivetti Research Laboratory (ORL)	40	400	p	(SAMARIA; HARTER, 1994)	17.55
Yale B	10	5,760	p,i	(GEORGHIADES; BELHUMEUR; KRIEGMAN, 2001)	13.83
Carnegie Mellon University pose, illumination and expression (CMU-PIE)	68	4,136	p,i,e	(GROSS et al., 2010)	10.64
University of Manchester Institute of Science and Technology (UMIST)	20	564	p	(WECHSLER et al., 2012)	6.38
Pointing Head and Pose Image Database (PHPD)	15	2,790	p,o	(GOURIER; HALL; CROWLEY, 2004)	6.38
Yale A	15	165	i,e	(BELHUMEUR; HESPANHA; KRIEGMAN, 1997)	4.25
Fédération Equestre Internationale (FEI)	200	2,800	p	(THOMAZ; GIRALDI, 2010)	3.72
Labeled Faces in the Wild (LFW)	1680	13,000+	p,i,e,o	(HUANG et al., 2008)	2.66
Aleix Martinez and Robert Benavente (AR)	126	4,000	i,e,o	(MARTINEZ, 1998)	2.13
Indian Face Database (IFD)	40	440	e	(JAIN, 2002)	1.59
Georgia Tech (GT)	50	750	p,i,e	(CHEN; MAN; NEFIAN, 2005)	1.59
Academy of Sciences pose, expression, accessories, and lighting (CAS-PEAL)	1,040	30,900	p,e,i,o	(GAO et al., 2008)	1.06
Japanese Female Facial Expression (JAFFE)	10	213	e	(LYONS et al., 1998)	1.06
Massachusetts Institute of Technology - Center for Biological and Computational Learning (MIT-CBCL)	10	2,000	p,i	(WEYRAUCH et al., 2004)	1.06
University of Essex	395	7,900	p,i,e,o	(SPACEK, 2002)	1.06
BIO-INFO	138	276	i	PRODOSSIMO et al.	0.53
BioID	23	1,521	p,e,o	(JESORSKY; KIRCHBERG; FRISCHHOLZ, 2001)	0.53
Intelligent Control & Computational Intelligence Laboratory (IC&CI)	15	825	p	(KIM; OH; KIM, 2016)	0.53
Korean Face Database	40	600	e,i	(TRINH; KIM; NA, 2014)	0.53
Poznan University of Technology (PUT)	100	9,971	p,e,o	(KASINSKI; FLOREK; SCHMIDT, 2008)	0.53
Honda/UCSD	20	560	p	(KIM; OH; KIM, 2016)	0.53
Vikrant Institute of Technology & Management (VITM)	-	200	a	(KHAN; GUPTA, 2016)	0.53
FG-Net	82	1,002	a	(KHAN; GUPTA, 2016)	0.53
IIIDelphi	102	2,618	p,i,e	(YADAV et al., 2013)	0.53

Variations are indicated by abbreviations: Expression (e), Illumination (i), Occlusion (o), Pose (p) and Aging (a)

Source: Own authorship

properties, we selected the techniques introduced in Section 2.1.1 and 2.1.2 to implement in our FR system. Regarding the optimization algorithms, PSO stood out among others revealing a reliable choice for future works, thus we chose this SI algorithm to use in our system. For comparison purposes, one of the most used EC algorithm was also employed, namely DE.

Following, some related works are presented in which DE and PSO are employed in FR systems as optimization algorithms.

3.3 RELATED WORK EMPLOYING DE AND PSO FOR FR

In our research for 2D FR systems employing DE, four works were found in the literature for feature selection. But, among these, three works were also employed for parameters optimization. Mallipeddi and Lee (2012) used the DE algorithm in their system to select the optimal Principal Component Analysis (PCA) features. In DE, each population member encodes the features to be selected from the total amount of features, and the search is guided by maximizing the distance between classes. To validate their methodology, Yale, Yale B, Olivetti Research Laboratory (ORL), and Aleix Martinez and Robert Benavente (AR) databases were used. Oh, Yoo and Pedrycz (2013) used DE to optimize the parameters of a Radial Basis Function Neural Network (RBFNN) and also for feature selection of combined PCA and Linear Discriminant Analysis (LDA) features. The DE is represented as a vector containing the learning rate, momentum coefficient, and fuzzification coefficient parameters, as well as the selected feature subset. In a later experiment, they used only the 2D-LDA features on the Yale A and ORL databases (YOO; OH; PEDRYCZ, 2013). Recently, they performed a comparative study of feature extraction methods and their application to RBFNN using the same FR systems architecture but with $2D^2$LDA features (OH; YOO; PEDRYCZ, 2016).

For PSO algorithm, most works found have used its binary version (BPSO) for feature selection, guiding the search through inter-class maximization and intra-class minimization distances on Discrete Wavelet Transform (DWT) feature space. Beside classical implementations of DWT (SURABHI et al., 2012; YAJI et al., 2012; RINKY et al., 2012; ROOHI et al., 2013; PRABHU et al., 2013; NIVEDITHA et al., 2015), other versions were employed, such as Threshold based DWT (VIDYA et al., 2012; MURTHY et al., 2012), Raster Scan Discrete Wavelet Transform (RDWT) (KHASHU et al., 2014) and a hybrid approach using Thresholded Wavelet Edge Enhancement Transform (TWEET) and DWT (HEGADE et al., 2013). Methods employing modified versions of BPSO were proposed, such as Threshold Binary Particle Swarm Optimization (ThBPSO) (KRISSHNA et al., 2014). In ThBPSO, the recurrence of selected features is considered based on a threshold set by the user. Sattiraju et al. (2013) proposed to use a modified version of BPSO called Adaptive Multi-Level Threshold Binary Particle Swarm Optimization (AMBPSO) to select features extracted by DWT. In AMBPSO, the 10 best global solutions are stored in a matrix and are used to guide the search. In addition to the works mentioned previously, Khadhraoui et al. (2016) used a deterministic parameter control technique on BPSO based on the number of iterations differing from traditional implementations. The benchmark databases mostly used to validate these methodologies are presented in descending order: Color Facial Recognition Technology (FERET), Yale B, ORL, Carnegie Mellon University pose, illumination and expression (CMU-PIE), University of Manchester Institute of Science and Technology (UMIST), Pointing Head and Pose Image Database (PHPD) and Fédération Equestre Internationale (FEI) database.

A couple of works used Discrete Cosine Transform (DCT) for feature extraction, and BPSO for selecting the most discriminant features based on inter-class maximization and intra-class minimization distances (DIVYA et al., 2012; GAGAN et al., 2012), moreover an accelerated version of BPSO (ABPSO) was employed on DCT feature space (ANEESH; MASAND; MANIKANTAN, 2012). In ABPSO, the velocity is updated for each iteration by summing it with the previous positional values for each particle. Ensemble approaches were experimented with DCT, such as Deepa et al. (2012) that used BPSO to select features extracted by Discrete Fourier Transform (DFT) and DCT, and Ajaya et al. (2014) proposed an approach to select extracted features from Contourlet Transform (CT) and DCT. On CT feature space, PCA was employed for dimensionality reduction and BPSO for feature selection. Last, Darestani, Sheikhan and Khademi (2013) extracted features using only CT and employed BPSO for feature selection. Then, in the classification stage, a classifier based on a artificial neural network was used with PSO-optimized hidden layer size and learning rate. In the works mention here, the mostly used databases are presented in descending order: ORL, Color FERET, Yale B, UMIST, CMU-PIE, PHPD, Japanese Female Facial

Expression (JAFFE) and Academy of Sciences pose, expression, accessories, and lighting (CAS-PEAL).

A considerable number of works using DWT and DCT along with PSO were also performed (D'CUNHA et al., 2013; NISCHAL et al., 2013; RAO et al., 2014; BABU et al., 2014; KODANDARAM et al., 2015). Also, Soumya et al. (2013) proposed a approach using BPSO to select features extracted from DWT and DCT. However, besides class separation, the authors used DCT trace in the fitness relation. Similarly Rao and Rao (2016) used BPSO for features selection extracted from DWT and Slope-form Triangular Discrete Cosine Transform (STDCT). In this work, fitness function was based in preserving maximum precision in order to represent the original feature set. The works mentioned in this paragraph used the following databases which are in descending order: Color FERET, CMU-PIE, PHPD, Yale B, UMIST, ORL, FEI, Georgia Tech (GT), Indian Face Database (IFD) and JAFFE.

Other approaches for features selection using BPSO and distance between classes were also proposed, such as the work developed by (SHANBHAG et al., 2014). They developed the work based on BPSO to search the features extracted by Wavelet Transform Feature Extraction (WTFE) for the optimal subset. Shetty et al. (2013) proposed a similar work, however, to select features extracted by Stationary Wavelet Transform (SWT) based technique namely Shift Invariance based Feature Extraction (SIFE). The authors claimed that they have used a modified version of BPSO called Weighted Binary Particle Swarm Optimizer (WBPSO), but, actually, they just considered the number of times a particular feature is being selected in the fitness computation. Also, Babu, Birajdhar and Tambad (2014) used the recurrence of a feature in fitness evaluation and called Conservative BPSO to select features from SWT feature space. Abhishree et al. (2015) proposed to use BPSO to select features extracted by Gabor filter technique. The same way, Kishore et al. (2014) used Gabor filter, but combined with Fast Fourier Transform (FFT). Differently, Cheng et al. (2014) searched for the optimal subset on the Self Quotient Image (SQI) features space meanwhile Nema and Thakur (2015) proposed to select features extracted by LDA. In their work, the algorithm was adapted with a deterministic parameter control technique which decreases c_1 and increases c_2 exponentially with time. Like others, the authors Vora et al. (2014) tried the features selection on Gamma Ray Burst Rhombus Star (GRBRS) feature mask space and Varun et al. (2015) on Block-wise Hough Transform (HT) feature space. The final work which used the distance between class as a fitness evaluation was proposed by Varadarajan et al. (2015). In this work, the author tested a modified version of BPSO called Exponential Binary Particle Swarm Optimization (EBPSO) to select the features extracted using Block based Additive Fusion. The group of articles mentioned in the present paragraph used the databases: Color FERET, CMU-PIE, Yale

B, PHPD, FEI, ORL, IFD, LFW, GT and CAS-PEAL.

Different approaches for feature selection using PSO and its variants were employed, such as Lei et al. (2012) who proposed to use a variation of PSO denominated as Fast Static Particle Swarm Optimization (FSPSO). It treats the whole initial feature set as a static particle swarm in which no new particle would be generated in high dimensional space, and the proposed method takes filter and wrapper strategy to pick out the most discriminative feature particle subset. In the universe of features selection works, the authors Shieh et al. (2014) proposed to use PSO to select features extracted by PCA using SVM as fitness function. Yin et al. (2014) proposed to use BPSO for feature selection on DCT coefficients feature space. In this approach, the fitness function is based on classification rate and dimensionality. In another work, YIN, FU and SUN (2014) used the same approach, but, they reduced the search space for BPSO algorithm by preselecting DCT coefficients according to a separability criterion. In addition to the modified versions of PSO, it is becoming common the use of modified versions BPSO. As an example, we can mention the work of the authors Sah et al. (2015) who proposed a modified version of BPSO called Logarithmic Binary Particle Swarm Optimization (LBPSO) to select features extracted by Entropic Gabor Wavelet Transform (GWT). In LBPSO, the global solution and particle best positions are weightage rather than their current positions. The feature vector is optimized by sampling the vector and choosing the one with the highest entropy. Among the works that we have mentioned, the work proposed by Mollaee and Moattar (2016) is somehow different since they proposed to use the PSO algorithm to aid discriminant ICA in finding multivariate data with lower dimension and independent features by maximizing Negentropy, as well as Fisher criterion. Zhang and Peng (2016) proposed to use PSO to find out the optimum combination of the basis and variety from a basis plus variety model on a high-dimensional unit sphere in terms of the minimum Euclidean distance relative to the query image. In their model, the query image is approximately a linear combination of the basis image and variety image. The basis images are the neutral images of subjects and variety images are generalized from multi-sample subjects. The basis of the optimum point gives the identity information for classification. The databases used by the works mentioned in this paragraph were: Yale B, ORL, Color FERET, CMU-PIE, AR and Massachusetts Institute of Technology - Center for Biological and Computational Learning (MIT-CBCL).

The bio-inspired approaches for parameters optimization were also employed the PSO algorithms. Here, we comment about the works related to SVM parameters optimization. The authors Valuvanathorn, Nitsuwat and Huang (2012) developed the work to aid SVM classification by selecting the parameters G and C automatically (PSO-SVM). In the next work, a modified version of PSO, called Opposition Particle

Swarm Optimization (OPSO) was presented by Hasan, Abdullah and Othman (2013) to optimize the SVM parameters in training and testing features extracted by PCA. In OPSO, two populations are generated: the first one is random and the second is opposition population which is based on the first population values. In this work, similar to others, the authors Xiao et al. (2014) used PSO combined with grid-search to optimize the parameters of a radial basis function kernel in SVM. Similarly, the authors Zou and Zhang (2016) presented their parameters optimization work using the recognition rate to calculate the fitness of each particle. The above mentioned works in the present paragraph used mostly the following databases: ORL, Color FERET, Yale A and BioID.

Besides optimizing SVM parameters, some works that optimize other parameters were also found, such as Pan, Zhu and Xia (2013) who proposed to replace exhaustive search used in Adaboost framework with PSO. The authors claimed that PSO is used as features selection procedure, but as each particle is encoded with a parameter set we considered as parameter optimization. The fitness function was defined as the normalized classification error rate. Banerjee and Datta (2013) proposed to use PSO for parameters optimization for both constrained and unconstrained type in which particle vectors are considered as correlator parameters to be optimized. The false acceptance rate was used as objective function to be minimized. Trinh, Kim and Na (2014) proposed to use PSO algorithm to find optimal weights to fuse global and local Fourier-Mellin Transform (FMT) features at score-level. The fitness function is evaluated by calculating the recognition rate related to a specific set of weights. Fernández-Martínez and Cernea (2015) proposed to use a modified version of PSO called Regressive Particle Swarm Optimization (RR-PSO) to optimize the parameters of SCAV1. SCAV1 is a supervised ensemble learning algorithm based on six nearest-neighbor classifiers based on histogram, variogram, texture analysis, edges, DWT and Zernike moments. In RR-PSO, global and local search are balanced by adopting regressive discretization in acceleration and in velocity of the PSO continuous model. Farag, Elghazaly and Hefny (2016) proposed to apply PSO to search the optimum HMM states and parameters. In their approach, maximum accuracy and minimum feature dimension was used to guide the search. Kim, Oh and Kim (2016) proposed to use PSO algorithm to optimize the parameters of RBFNN such as the number of nodes and fuzzification coefficient. The classification rate is used as the fitness value. Finally, Plichoski, Parpinelli and Chidambaram (2017) proposed to use PSO to optimize Homomorphic filter (HF) parameters namely high and low frequency factors, cut-off frequency and filter's order. The recognition rate is used as fitness function to guide the algorithms search. The most used database by the presented works were: ORL, CMU-PIE, Yale A, PHPD, MIT-CBCL, AR, Korean Face Database, Poznan University of Technology (PUT), Honda/UCSD and Intelligent Control & Computational Intelligence Laboratory (IC&CI).

From related works, it is possible to find that DE or PSO are the most used optimization algorithms for feature selection and/or parameters optimization. In feature selection procedure, the features extracted from face images are encoded and the optimization algorithm attempts to maximize inter-class distances selecting a subset from the original full set. For parameters optimization, distinct techniques are proposed and their parameters are encoded in the optimization algorithm which attempts to optimize an objective function, searching for an optimized configuration that leads to better recognition rates. All these approaches are suitable for FR systems since they provide a trade-off between execution time and precision. Inspired by the works that employed optimization algorithms for parameters optimization, it is proposed here an approach for FR which optimizes not only all the parameters involved but also the techniques that will be used. Hence, it is possible to group several distinct techniques that are proposed in literature, and let the optimization algorithm search for the best settings.

4 RELATED WORK

Over the years, much interest and research have been focused on the field of FR and consequently, an increased number of methodologies have been proposed. In this section, we present the related work, which will also be used for further comparison in Section 6.2 and 6.3.

Among the aforementioned challenges found in FR systems, uncontrolled illumination is a critical factor for real-world applications due to the non-ideal imaging environments frequently encountered in practical situations (ZHANG et al., 2018). In the review of illumination compensation approaches performed in 2015 (OCHOA-VILLEGAS et al., 2015), some authors presented their works with most significant results. Firstly, Cheng, Li and Jiao (2013) proposed the use of Discrete Wavelet Transform (DWT) technique for illumination normalization based on retinal modeling executed on low frequency band. In Baradarani, Wu and Ahmadi (2013), it is shown that multi-scale analysis of facial structure and features of face images leads to high recognition rates for images under varying illumination. They first assume that an image is a black box consisting of a combination of illumination and reflectance. As illumination resides in the low-frequency part of image, they worked with a high-performance multi-resolution transformation to separate the frequency components of an image, namely Double-Density Dual-Tree Complex Wavelet Transform (DD-DTCWT). After extracting a mask, feature vector is formed and the PCA is used for dimensionality reduction which is then proceeded by the Extreme Learning Machine (ELM) as a classifier to evaluate the performance of the proposed algorithm for face recognition under varying illumination. After, Baradarani and Wu (2013) also proposed the use of resonance based decomposition of images for illumination invariant face recognition. Although illumination is mostly considered as the low-frequency part of images, these low-frequency contents may consist of low- and/or high-resonance nature. They assume that an input image can be considered as a combination of illumination and reflectance, this way the energy distribution are different for an image with good illumination effects and an image with high illumination variations. Based on this assumption, the energy of sub-bands of the two components are thresholded to deactivate the sub-bands with unwanted energy distribution created by illumination effects. Once again, PCA and ELM were used for dimensionality reduction and classification, respectively. In Vishwakarma (2015), the authors believed that the effect of illumination variations is in decreasing order over low-frequency DCT coefficients. Hence, they proposed an approach to nullify the effect of illumination variations as well as to preserve the low-frequency details of a face image using a fuzzy filter. Hu (2011) proposes a DWT-based denoising technique to detect the illumination discontinuities in the detail sub-bands. And the detail coefficients are updated with the use of the obtained discontinuity information. Then, a smooth version of the input image is obtained by applying the inverse DWT on the updated wavelet coefficients. Finally, a multi-scale reflectance model is presented to extract the illumination invariant features. Lastly, Zhao et al. (2013) proposes an approach for face recognition based on Grayscale Arranging Pairs (GAP). A facial model is built by using the stable point pairs of the GAP features. Then the similarity between the facial model and the input facial image is calculated by checking whether the intensity relationship of these point pairs is the same.

Besides these, recent works still attempt to address the illumination related issues such as Yang et al. (2017) that presented a regression method for FR. Most existing regression methods use the one-dimensional, pixel-based error model, which characterizes the representation error individually, pixel by pixel, and thus neglects the two-dimensional structure of the error image. They observed that illumination changes to a low-rank error image. In order to make use of this low-rank structural information, their paper presents a 2D image-matrix-based error model, namely, Nuclear norm based Matrix Regression (NMR). The NMR uses the minimal nuclear norm of representation error image as a criterion, and the Alternating Direction Method of Multipliers (ADMM) to calculate the regression coefficients. The authors show that their method has overcome the state-of-the-art regression-based methods. In addition to these works, Varadarajan et al. (2015) also proposed a FR system using a combination of Chirp Z-Transform (CZT) and Goertzel algorithm for illumination normalization and Binary PSO optimization algorithm for feature selection. This work is covered in the research performed in Section 3.3.

The OSPP problem is a real-world condition and some works that attempt to solve this problem can be highlighted. In Hu et al. (2017) 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 (R&L) and the High-and Low-frequency (H&L). 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. Gu, Hu and Li (2018) 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 Shang et al. (2018), 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.

Some works found in the literature use neural networks to deal with the OSPP problem. Guo et al. (2017) proposed a two-layer 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. Oh et al. (2018) proposed a single hidden layer analytic Gabor feedforward 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.

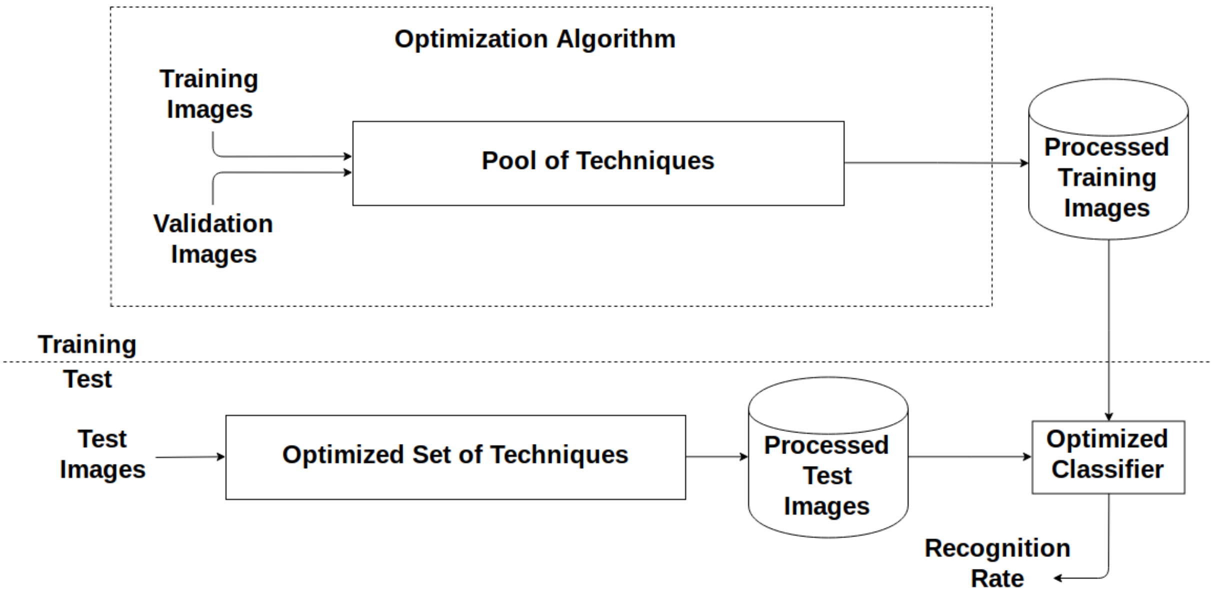
Finally, some distinct methodologies are also presented. In Mehrasa, Ali and Homayun (2017) the authors presented a supervised learning method called supervised locality preserving multimanifold (SLPMM). In their approach, two graphs are made to represent the information inside every manifold and among different manifolds, with this it simultaneously maximizes the between-manifold scatter and minimizes the within-manifold scatter. Another work considering the OSPP scenario is proposed by Chu, Zhao and Ahmad (2018), in which the authors 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.

5 THE FR FRAMEWORK FOR ILLUMINATION COMPENSATION

This section details each step of the proposed approach. The FR framework provides techniques for preprocessing and for 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. Besides that, the parameters of selected techniques are also optimized by the framework as well as the distance metric employed by the classifier.

Figure 8 represents the proposed FR system's flowchart. The dataset should be split into three subsets: training, validation and test. In training stage, the validation subset is used to evaluate the candidate solutions generated by the optimization algorithm based on the training subset. In test stage, the test subset is used to evaluate the optimized model generated after training. In both training and test stages the fitness evaluation is done based on the recognition rate.

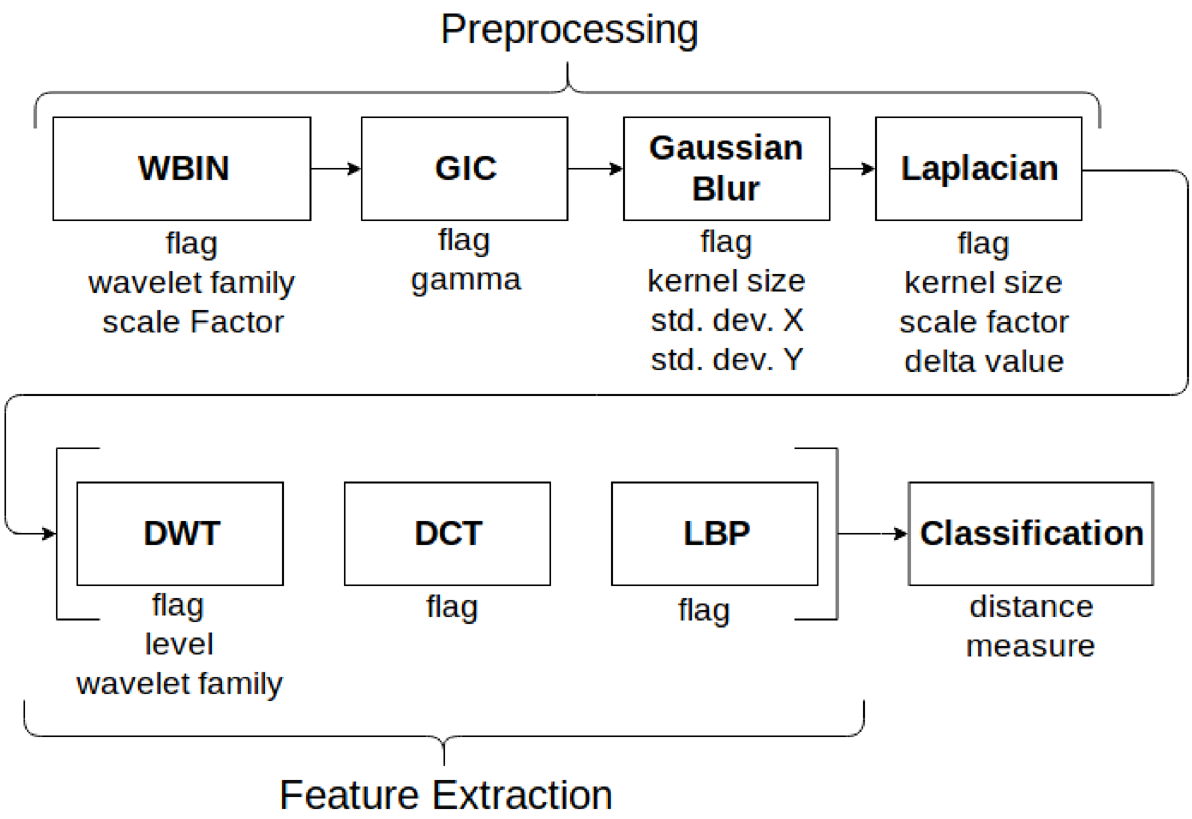
Figure 8 – Proposed FR systems flowchart.



Source: Own authorship

In this work, we implemented the techniques introduced in Section 2.1.1 and 2.1.2, as well as the distance measures presented in Section 2.1.3. For preprocessing, the first technique employed was WBIN to adjust contrast and enhance face structure. After, the Gamma Intensity Correction (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 only if the techniques are applied and its parameters but not the order they are 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 9 illustrates the employed techniques and their respective parameters.

Figure 9 – Pool of Techniques employed in the FR flowchart.



Source: Own authorship

The optimization algorithm acts in training stage and the techniques as well as their parameters are encoded in the solution vector as real-coded variables. Figure 10 shows in (a) how the solution vector is encoded and in (b) a random sampling of a possible instance. The solution vector consists of the WBIN, the GIC, the Gaussian Blur, the Laplacian Edge Detection, the DWT, the DCT, the 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 operates in continuous domain, the parameter $flag$ is discretized to 0 or 1. Also, due to the stochastic nature of the optimization process,

Figure 10 – Solution encoding and a possible configuration.

(a) Solution encoding

flag	wavelet family	scale factor	flag	gamma	flag	kernel size	std. dev. X	std. dev. Y

flag	kernel size	scale factor	delta value	flag	level	wavelet family	flag	flag	distance measure

(b) Random sampling of a possible solution

0	1	2	3	4	5	6	7	8
0.85	41	1.25	0.01	1.5997	0	30	1.1987	4.0293

9	10	11	12	13	14	15	16	17	18
0.42	7	2.3966	1.8881	0.47	3.32	10	0.29	0.54	3

Source: Own authorship

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 our proposed approach.

The pseudo-code for fitness evaluation is shown in Algorithm 5. As input, the algorithm receives a set of parameters ($\vec{p}$) representing a possible solution (as show in Figure 10) and the images' path ($imgs\_search,\ imgs\_ref$) for classification. The output is the recognition rate ($accuracy$) employing the referred techniques. This algorithm is used to evaluate the model, since both codes consists in applying the techniques on the validation (or test) images ($imgs\_search$) and on the training images ($imgs\_ref$) based on the solution vector $\vec{p}$, and returning the recognition rate ($accuracy$). In lines 1 and 2, the images are $read$ from their respective paths. In lines 3 to 30, the techniques are applied based on the input parameters using their specific functions ($wbin$, gic, $gaussian\_blur$, $laplacian$, dwt, dct and lbp). It is worth to point out that the indices representing the $flag$ parameters ($0,\ 3,\ 5,\ 9,\ 13,\ 16\ and\ 17$) are discretized inside the fitness/evaluation function with the threshold of 0.5. In line 31, the distance measure employed is mapped to a variable. Lines 32 to 45 show the procedure of comparing all search images to every reference image, and returning the most similar reference image for each search image. Finally, the recognition rate is calculated and returned as output.

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 stop criterion is reached, the setting that achieved the best recognition rate is selected as 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.

Algorithm 5 The pseudo-code for fitness evaluation

```
Input: p⃗, path_ref, path_search
Output: accuracy
imgs_ref ← read(path_ref)
imgs_search ← read(path_search)
if p[0] ≥ 0.5 then
    imgs_ref ← wbin(imgs_ref, p[1], p[2])
    imgs_search ← wbin(imgs_search, p[1], p[2])
end if
if p[3] ≥ 0.5 then
    imgs_ref ← gic(imgs_ref, p[4])
    imgs_search ← gic(imgs_search, p[4])
end if
if p[5] ≥ 0.5 then
    imgs_ref ← gaussian_blur(imgs_ref, p[6], p[7], p[8])
    imgs_search ← gaussian_blur(imgs_search, p[6], p[7], p[8])
end if
if p[9] ≥ 0.5 then
    imgs_ref ← laplacian(imgs_ref, p[10], p[11], p[12])
    imgs_search ← laplacian(imgs_search, p[10], p[11], p[12])
end if
if p[13] ≥ 0.5 then
    imgs_ref ← dwt(imgs_ref, p[14], p[15])
    imgs_search ← dwt(imgs_search, p[14], p[15])
end if
if p[16] ≥ 0.5 then
    imgs_ref ← dct(imgs_ref)
    imgs_search ← dct(imgs_search)
end if
if p[17] ≥ 0.5 then
    imgs_ref ← lbp(imgs_ref)
    imgs_search ← lbp(imgs_search)
end if
distanceMeasure = p[18]
matches = 0
for search = 0 to imgs_search's size do
    minDist = ∞
    for ref = 0 to imgs_ref's size do
        dist = calcDist(imgs_search[search], imgs_ref[ref], distanceMeasure)
        if dist ≤ minDist then
            minDist = dist
            most_similar_image = imgs_ref[ref]
        end if
    end for
    if imgs_ref's label == most_similar_image's label then
        matches = matches + 1
    end if
end for
accuracy = (matches/imgs_search's size) * 100
return accuracy
```

6 EXPERIMENTS AND RESULTS

In this chapter, we present the design of experiments as well as the results obtained. A fair discussion is performed about the optimization algorithms employed, presenting comparison charts i.e. box-plot of the experiments and statistical analysis. Also, a comparison with literature results is made.

6.1 DESIGN OF EXPERIMENTS

As DE and PSO 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 standard deviation of 2.5% (LILJA, 2005). Based on this criterion, 10 runs were performed for each algorithm in each dataset.

To the best of our knowledge, all works that attempt to design a FR system used the recognition rate or the number of errors as performance metric. In this dissertation, we used the recognition rate, which is the ratio between the number of reference face images assigned to the correct class by the total amount of reference face images.

Tables 2 and 3 shows the parameters for PSO and DE, respectively. The common parameters are shown in Table 4. It is worth remembering that the jDE algorithm does not require the definition of the control parameters (F and CR), however the common parameters are employed as in Table 4. All parameters were set using literature recommendations (BROWNLEE, 2011).

Table 2 – PSO parameters.

Particle Swarm Optimization	
w	0.9
c_1	2
c_2	2
V_{max}	10% of the size of the domain

Source: Own authorship

Table 3 – DE parameters

Differential Evolution	
F	0.8
CR	0.9

Source: Own authorship

Table 4 – Common parameters for both algorithms.

Common Parameters	
$Population\ Size$	20
$Stop\ Criteria$	100 iterations

Source: Own authorship

Statistical tests are required to ensure statistical significance in the results obtained. Based on this, we first applied the Shapiro and Wilk normality test (SHAPIRO; WILK, 1965) and, once verified, the Wilcoxon signed-rank paired test (WILCOXON,

1945) is applied. Both tests were performed considering a 95% confidence interval. These tests were chosen since they are the most commonly used in literature.

Three case studies are carried out in this work.

The first case study uses the Yale Extended B database and aims to compare the two canonical optimization algorithms (i.e. DE and PSO) considering the illumination variation challenge.

For the second case study, only the algorithm that turned out better in the first case study was employed (i.e. DE). Besides that, since it was verified that the bottleneck in the proposed approach remains in training step, specially in the classification of the validation images, a different configuration was employed for comparison purposes. Instead of using 50% for validation and test split, it was used 10% for the validation subset and 90% for the test subset. Also, it was defined as stop criterion to the system when reaching 100% of recognition rate. The same common parameters were employed, however, aiming to drop the diversity to a more local search, the parameter F was set to 0.5 instead of 0.8. This case study were performed on more challenging datasets, since the proposed approach almost addressed the Yale Extended B database in 100%. The datasets employed are: the CMU-PIE dataset that contains different poses and illumination variation when only a single training image per subject is available (OSPP problem), and the b-series subset from FERET database (PHILLIPS et al., 1998) that consists of images with facial pose angles ranging from +60° to –60°. Further details about the databases and their division are described on the following subsections.

In the third case study, a self-adaptive version of DE (jDE) is employed with the new settings proposed in the second case study, and compared with the DE canonical version. The goal here is to minimize the number of parameters to adjust and analyze the performance of the framework. For this case study, we also used the CMU-PIE and FERET database with the same experimental methodology employed in case study 2.

To evaluate the optimization algorithms on-the-fly, convergence and diversity charts were generated. The convergence chart represents the recognition rate of the best solution at each iteration from which we can visualize the algorithm's convergence. The population's diversity chart is obtained using Equation 6.1 (CORRIVEAU et al., 2013).

$$Diversity = \frac{\sum_{i=1}^{Pop_{size}-1} \ln\left(1 + \min\limits_{j\in[i+1,Pop_{size}]} \frac{1}{Prob_{size}} \sqrt{\sum_{k=1}^{Prob_{size}} (x_{i,k} - x_{j,k})^2}\right)}{NMDF} \tag{6.1}$$

Where, Pop_{size} is the population size, $Prob_{size}$ is the problem size and $NMDF$

is the maximum diversity obtained so far in the optimization process. This chart represents how spread the solutions are in the search space. With this it possible to draw whether the algorithm is performing a more local or global search at each iteration.

All experiments were run in Ubuntu 18.04.1 LTS with an Intel® Core i7-4770 computer and 16 GB RAM memory using Python programming language version 3.6.7 with multiprocessing library for parallelism. Parallelism is achieved distributing the fitness evaluation of optimization algorithms population in 7 out of 8 available processing cores. The libraries for image processing were OpenCV 3 and PyWavelets. Following, we describe the datasets used in this work.

6.1.1 Yale Extended B database

According to the research performed in Section 2.2, the Yale B database is a well known database because of its several degrees of illumination. Based on this, the Yale B extended version is used in this work. This database consists of 38 subjects under different conditions varying pose and illumination (9 poses x 64 illuminations). As our purpose is to compensate the illumination variation, only the frontal face images were selected. This database is commonly divided into five subsets according to the angle between the light source direction and camera axis (OCHOA-VILLEGAS et al., 2015). Table 5 shows the angle variations, as well as the number of subjects, and Figure 11 illustrates the sample images from subsets 1 to 5. The well-illuminated images from S1 are used as training set. The remaining subsets are divided in half as validation and test sets.

Table 5 – Five subsets according to the light angle source directions.

	S1	S2	S3	S4	S5
angle	0^o-12^o	13^o-25^o	26^o-50^o	51^o-77^o	>78^o
number of images	6	12	11	12	16

Source: Own authorship

6.1.2 CMU-PIE database

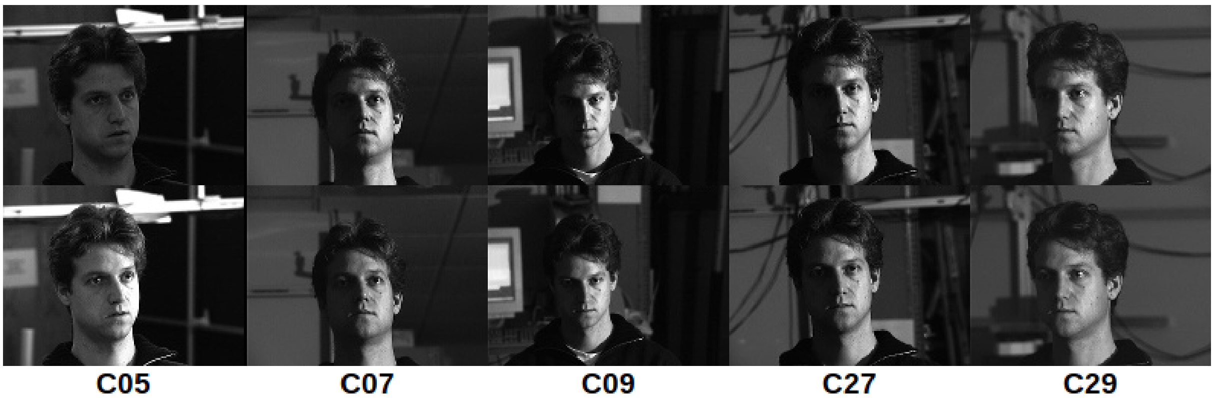
As shown in Table 1 from Chapter 2, the CMU-PIE (SIM; BAKER; BSAT, 2002) database consists of 41,368 images of 68 subjects. Each subject has several images under 13 different poses and 43 illumination conditions, and with 4 expressions. For case study 2, it was selected five subsets based on the camera used (i.e. C05, C07, C09, C27 and C29). Each camera was set in different angles, this way presenting five different poses: 1) looking right; 2) up; 3) down; 4) front; and 5) left, respectively. Figure 12 illustrates the five subsets. For a fair comparison with literature, we have selected about 24 images per subject with different illumination conditions according

Figure 11 – Example of Yale Extended B database subsets from S1 to S5, respectively.

Source: Own authorship

to database documentation. Each one of the five subsets are evaluated independently and evaluated 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 and thus, 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 characterizes the OSPP problem.

Figure 12 – Example of CMU-PIE database subsets C05, C07, C09, C27 and C29, respectively.



Source: Own authorship

6.1.3 FERET

The b-series subset from FERET database (PHILLIPS et al., 1998) was employed, which consists of 1800 images from 200 persons with facial pose angles ranging from +60° to –60° and size of 256x384. Table 6 presents the description of subsets used in this experiment and Figure 13 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 6 – 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 13 – Example of FERET b-series subsets.

Source: Own authorship

6.2 CASE STUDY 1

In this section, the experimental results of the proposed methodology using PSO and DE algorithms with the Yale Extended B database are exposed and discussed.

Table 7 shows the recognition rates achieved in training and test stages for both optimization algorithms, as well as their respective average and standard deviation. For both optimization algorithms, the high recognition rates demonstrates the efficiency in finding an optimized solution, as well as the performance of the chosen techniques in compensating illumination variation. Similar results are obtained in both stages, which means that precision on training reflects well on test, this way, preventing overfitting. In addition to this, the low standard deviation values implies the low sensitiveness to random variables in an optimization approach. According to the similar recognition rates obtained from the different sets of techniques, we can conclude that the problem space is highly multi-modal, i.e. where different solution vectors lead to same or almost the same fitness values.

Using the confidence interval concept is possible to determine how many measurements are necessary to produce reliable results (LILJA, 2005). With this, it was found that 10 experiments are enough to analyze the proposed methodology considering a 95% confidence interval. Regarding the performance of the algorithms to find

Table 7 – Recognition rates (%) for each experiment - case study 1

	Particle Swarm Optimization		Differential Evolution	
Experiment	**Training**	**Test**	**Training**	**Test**
1	98.38	98.49	99.80	99.80
2	99.61	99.23	99.80	99.58
3	99.69	99.80	99.91	99.23
4	99.69	99.67	99.78	99.72
5	99.69	99.47	99.80	99.61
6	99.72	99.72	99.91	99.50
7	98.98	98.49	99.89	99.61
8	99.80	99.72	99.91	100
9	99.58	99.30	99.80	99.52
10	98.35	97.86	99.69	99.91
Average	99.34 $\pm$ 0.56	99.17 $\pm$ 0.66	99.83 $\pm$ 0.07	99.65 $\pm$ 0.22

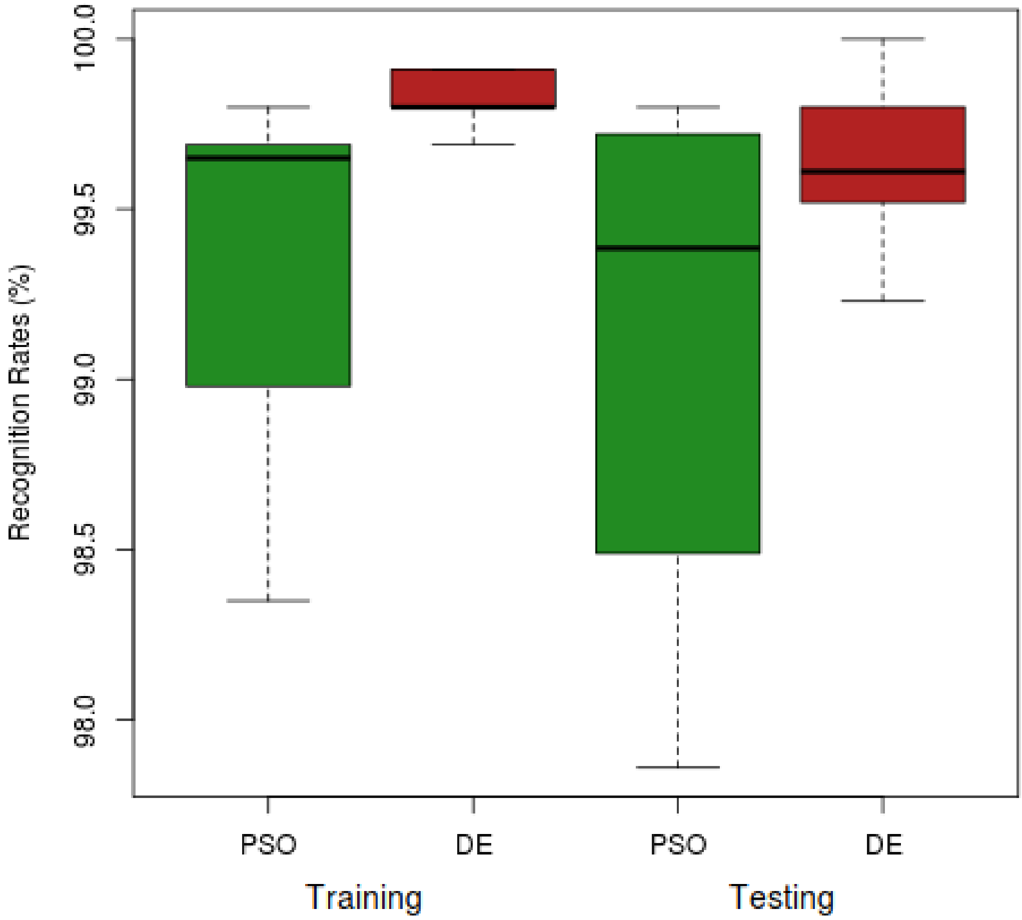
Source: Own authorship

an optimized solution, the box-plots for both training and test stages are presented in Figure 14. The DE presented more accurate results, achieving high recognition rates in most experiments, compared to PSO in both training and test. However, we cannot assert that a significant difference between these two algorithms exist and consequently, it may demand additional statistical tests.

For this, the Shapiro-Wilk normality test is applied in the results obtained by both PSO and DE (SHAPIRO; WILK, 1965). From the results, for both algorithms, the hypothesis that the samples comes from a normal distribution is rejected with a confidence level of 5% (i.e. $p - value < 0.05$). Hence, the non-parametric Wilcoxon signed-rank paired test (WILCOXON, 1945) is applied to verify the statistical significance of the results obtained by PSO and DE, also considering a confidence level of 5%. Since the $p - value = 0.0058$ from training stage it is possible to conclude that the results obtained by the DE algorithm are statistically better than the results obtained by the PSO algorithm. On the other side, with the $p - value = 0.0923$ from test stage it is possible to conclude that the results obtained by both algorithms are statistically the same.

Regarding processing time, the Shapiro and Wilk test rejects the hypothesis that the samples comes from a normal distribution, this way, in this case we also employed the Wilcoxon signed-rank paired test which rejected the hypothesis of significant difference ($p - value > 0.05$). The average training time was 12 hours and 49 minutes with a standard deviation of 10 hours and 38 minutes for PSO, and 5 hours and 48 minutes with standard deviation of 1 hour and 11 minutes for DE. However, after the training task is completed, with the definition of the optimized settings, the average time to test an image using PSO is 0.2344 seconds with 0.2479 standard deviation and using DE, 0.0731 seconds with 0.0130 standard deviation. In Figure 15, the box-plot for training and test stages is presented. It is possible to draw the same conclusion that we

Figure 14 – Box-plot for both algorithms performance in training and test stages - case study 1

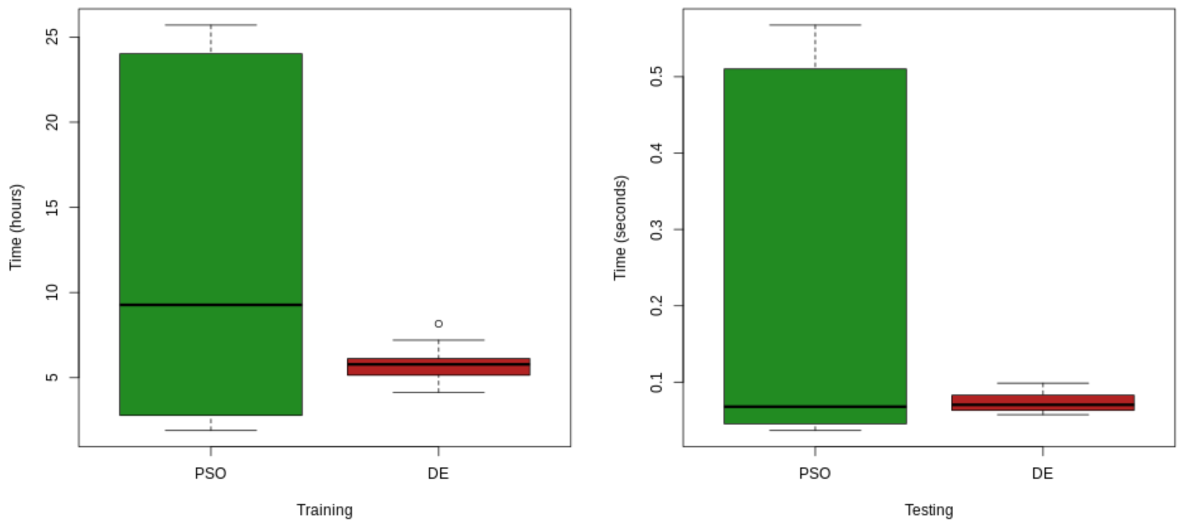


Source: Own authorship

obtained when comparing performance, which is that DE presented a lower standard deviation and thus, more consistent results.

Following, Table 8 presents the settings found by PSO algorithm in the optimization step. Each row represents the configuration obtained in each experiment, in the same order as shown in Table 7. Each column represents a parameter of the employed techniques. A dash symbol means that the related technique was not chosen. It is possible to observe how distinctive they are from each experiment. Among the proposed techniques, the most used ones in the training stage were GIC, Laplacian, DWT and DCT. Table 9 presents the techniques chosen by the DE optimization algorithm for each experiment. From the experimental analysis based on the optimization approach,

Figure 15 – Box-plot for both algorithms time in training and test stages - case study 1



Source: Own authorship

the GIC and Laplacian techniques were chosen by the algorithm in every final solution. This indicates that the selected combination of techniques are more consistent than the ones found by PSO. In most of the experiments, the L_{2sqr} distance measure was chosen, although same results are accomplished with L_1. In addition to this, the LBP texture descriptor was never chosen.

Table 8 – Settings from training step using Particle Swarm Optimization - case study 1

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	WF	SF	γ	$Kernel$	sdX	sdY	$Kernel$	SF	$Delta$	WF	$level$			
1	-	-	2.5108	3	2.5481	1.6057	-	-	-	-	-	-	Yes	L_1
2	-	-	-	-	-	-	13	2.8910	2.8857	coif13	3	Yes	-	L_{2sqr}
3	-	-	3	-	-	-	13	2.1997	1	bior1.1	1	Yes	-	L_2
4	-	-	3	-	-	-	15	1.9376	1	db32	3	-	-	L_2
5	-	-	-	-	-	-	13	2.1018	1	coif17	3	Yes	-	L_2
6	-	-	3	31	1	3.3355	13	3	2.6992	coif8	3	-	-	L_{2sqr}
7	sym4	0	3	-	-	-	-	-	-	-	-	-	yes	L_1
8	-	-	3	15	1.4697	4.1963	7	1.7236	2.7023	-	-	Yes	-	L_2
9	sym20	1.0254	2.8635	31	2.1272	2.0744	3	2.8153	3	sym20	1	Yes	Yes	L_2
10	-	-	2.8672	27	1	1.8493	-	-	-	-	-	Yes	Yes	L_1

Source: Own authorship

Finally, Figures 16 and 17 presents the average convergence and average diversity charts of PSO and DE optimization tasks for all experiments, respectively. A similar behavior is observed in both convergence charts, close to the 40th iteration the best solution improves in a slow rate but continuously improving when the stop criterion is reached. However, in most experiments, almost maximum recognition rate is reported justifying 100 iterations as stop criteria. Due to the multi-modal search space

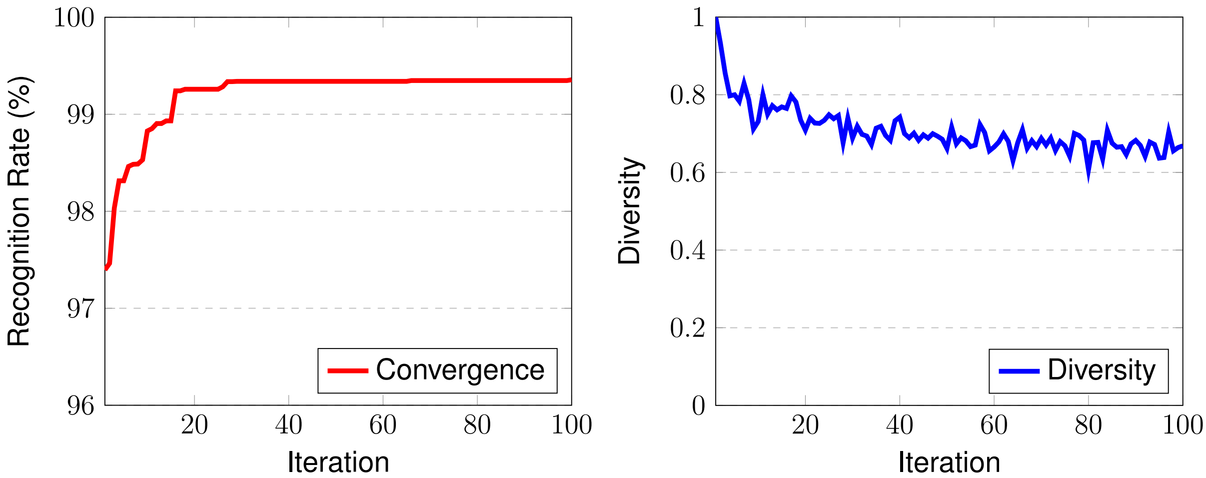
Table 9 – Settings from training step using Differential Evolution - case study 1

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	-	-	2.5108	5	5	5	7	2.7198	3	sym9	1	-	-	L_{2sqr}
2	-	-	3	-	-	-	13	1.3637	2.8603	coif12	2	-	-	L_{2sqr}
3	db5	1.2606	2.8723	3	1	3.5963	15	2.0067	1	-	-	Yes	-	L_{2sqr}
4	-	-	2.0171	-	-	-	15	1.8360	1.9107	sym10	1	-	-	L_{2sqr}
5	-	-	3	-	-	-	15	3	2.1105	-	-	-	-	L_{2sqr}
6	-	-	2.3487	9	1	3.3040	13	3	1	-	-	-	-	L_{2sqr}
7	-	-	1.7961	-	-	-	15	2.8252	3	-	-	Yes	-	L_{2sqr}
8	bior3.7	0.6483	2.2690	15	1.7642	5	7	1	3	bior1.1	1	Yes	-	L_1
9	bior1.1	0.0343	2.7519	3	5	1	7	1	1	-	-	-	-	L_1
10	-	-	2.9877	3	5	2.4727	15	1.2302	1.1656	-	-	-	-	L_2

Source: Own authorship

nature of the problem, the diversity in our experiments remained high during all optimization step for both algorithms.

Figure 16 – Convergence and Diversity charts for Particle Swarm Optimization optimization task, respectively - case study 1



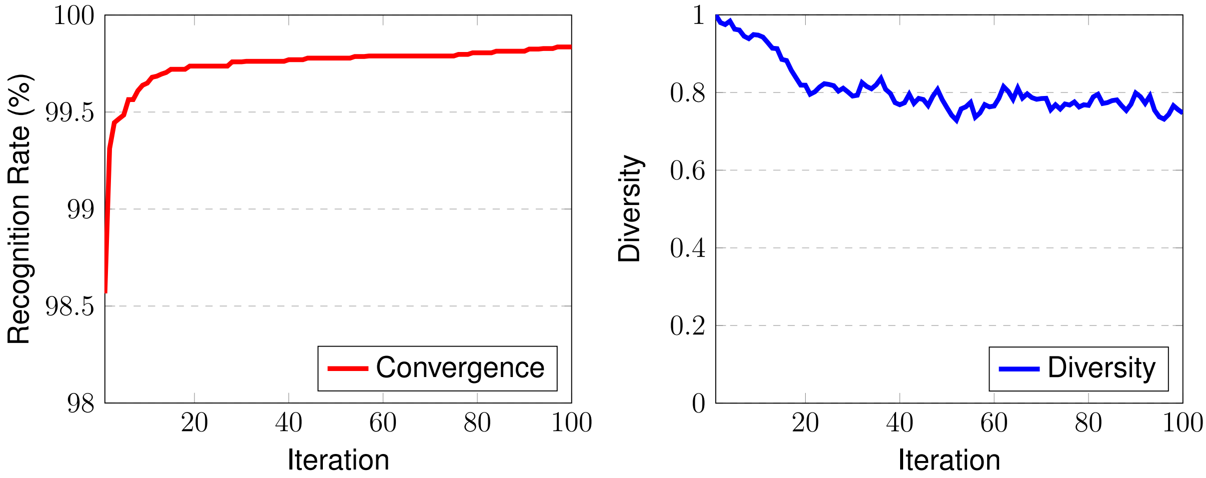
Source: Own authorship

In this experiment, it was possible to draw that DE algorithm is superior than PSO for this problem, achieving more consistent results considering both accuracy and processing time. Hence, we employed only DE algorithm in further experiments. Also, it was verified that the bottleneck remains in the classification of the validation images on training step, this way, responsible for the high processing time. Hence, a different configuration described in Section 6.1 is employed and compared in Case Study 2.

6.3 CASE STUDY 2

In this case study, we intend to compare a different configuration of the DE algorithm, and evaluate the proposed FR system on other databases to verify robustness (i.e. CMU-PIE and FERET). For convention purposes, we will refer to the configuration

Figure 17 – Convergence and Diversity charts for Differential Evolution optimization task, respectively - case study 1



Source: Own authorship

employed in the first case study as DE$_A$. The new configuration aims to reduce processing time allowing the algorithm to stop if 100% recognition rate is achieved on training step. Also, only 10% of the images are used for validation. For this configuration, we will refer to DE$_B$. Following, a discussion about the results is presented.

Table 10 presents the recognition rates using DE$_A$ and DE$_B$ on CMU-PIE subsets. The recognition rates for training and test stages are in the columns while the experiments are in rows. The last row represents the average recognition rate for all experiments on each subset. Regarding DE$_A$ (Table 10a), it is possible to observe that the recognition rate achieved a value higher than 99% in every case, proving that the proposed FR system has the ability to extract discriminant features from face images containing slightly changes in pose, being robust across different degrees of illumination. Furthermore, when pose and illumination variation are considered within the dataset, the proposed approach showed to be suitable since it achieved very good recognition rates. For subsets C05, C09 and C29, an accuracy of 100% was achieved in most of the experiments, while the accuracy for the other subsets were very close to 100%. Considering DE$_B$ (Table 10b), one can draw the same conclusions despite the reduction of in the number of validation images. The only exception is subset C27, in which a recognition drop of about 3% and 4% is verified for training and test stages, respectively. In addition, a lost in precision is observed, specially on subsets C09 and C29. Despite these inferences, the statistical tests employed reveled that we cannot reject the hypothesis of significant difference considering subsets C07 ($p-value = 0.0112$), C09 ($p-value = 0.0074$), C27 ($p-value = 0.0050$) and C$_{All}$ ($p-value = 0.0365$) on test and C27 ($p-value = 0.0050$) on training stage.

Table 10 – Recognition rates (%) for each experiment on CMU-PIE subsets - case study 2

(a) DE_A

	C05		C07		C09		C27		C29		C_{All}	
Experiment	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**
1	100	100	99.55	100	100	100	99.55	99.55	100	99.85	99.85	99.82
2	100	100	99.55	100	100	100	99.85	99.41	100	100	99.91	99.73
3	100	100	100	99.55	100	99.85	99.55	99.55	100	100	99.85	99.88
4	100	100	99.85	99.55	100	100	99.70	98.97	100	100	99.91	99.82
5	100	100	99.85	99.70	100	100	99.85	99.70	100	100	99.85	99.82
6	100	100	99.85	99.41	100	100	99.55	99.70	100	99.26	99.82	99.91
7	100	100	99.85	99.70	100	100	99.70	99.55	100	100	99.91	99.76
8	100	100	99.55	100	100	100	99.70	99.55	100	99.85	99.79	99.91
9	100	100	99.85	99.70	100	100	99.85	99.55	100	100	99.88	99.82
10	100	100	99.70	99.70	100	99.85	99.85	99.55	100	100	99.88	99.82
Average	100	100	99.76 $\pm$ 0.16	99.73 $\pm$ 0.20	100	99.97 $\pm$ 0.06	99.71 $\pm$ 0.13	99.5 $\pm$ 0.20	100	99.90 $\pm$ 0.23	99.86 $\pm$ 0.04	99.82 $\pm$ 0.05

(b) DE_B

	C05		C07		C09		C27		C29		C_{All}	
Experiment	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**	**Training**	**Test**
1	100	100	99.26	99.76	100	99.92	96.75	94.54	100	99.92	99.85	99.67
2	100	100	100	99.76	100	99.92	95.19	93.08	100	99.75	99.85	99.79
3	100	100	100	99.10	100	99.84	96.70	96.34	100	99.92	99.85	99.66
4	100	100	100	99.02	100	100	93.59	92.74	100	99.92	99.71	99.84
5	100	100	100	99.84	100	99.84	97.76	95.41	100	99.59	100	99.80
6	100	100	100	99.43	100	94.69	98.72	94.71	100	99.75	99.85	99.80
7	100	100	99.26	99.18	100	99.92	98.35	95.86	100	99.18	99.85	99.85
8	100	99.50	99.26	98.94	100	99.51	97.71	96.05	100	99.51	100	99.56
9	100	100	100	99.59	100	99.18	97.12	95.88	100	100	99.85	99.84
10	100	100	100	99.10	100	99.35	97.12	96.07	100	99.92	100	99.62
Average	100	99.95 $\pm$ 0.16	99.78 $\pm$ 0.36	99.37 $\pm$ 0.34	100	99.22 $\pm$ 1.61	96.90 $\pm$ 1.52	95.07 $\pm$ 1.28	100	99.75 $\pm$ 0.25	99.88 $\pm$ 0.09	99.74 $\pm$ 0.10

Source: Own authorship

Table 11 presents the results regarding FERET database. As we can see, substantial recognition rates were achieved taking into account that most techniques implemented in the framework deals only with illumination. It is clear that there is no statistical difference in training step. On the other hand, a considerably drop in recognition rate is verified in test step, thus presenting statistical difference ($p-value = 0.005$). With this, we conclude that reducing the number of images in the validation subset at this degree results in a substantial loss of accuracy in test stage.

Table 11 – Recognition rates (%) for each experiment using FERET database - case study 2

	DE_A		DE_B	
Experiment	**Training**	**Test**	**Training**	**Test**
1	79.25	81.75	84.38	79.72
2	80.00	79.50	81.88	79.17
3	79.87	80.62	76.88	71.88
4	80.87	80.00	78.12	78.75
5	80.62	79.75	80.62	77.78
6	81.25	79.12	83.12	78.40
7	79.75	80.75	82.50	77.29
8	79.87	81.12	78.75	78.33
9	80.62	79.87	78.75	75.69
10	80.37	79.37	82.50	78.06
Average	80.24 $\pm$ 0.60	80.18 $\pm$ 0.84	80.75 $\pm$ 2.50	77.51 $\pm$ 2.26

Source: Own authorship

Table 12 shows the average processing time for each subset in training stage and the average processing time to test a face image using DE$_A$ and DE$_B$ for CMU-PIE subset, followed by their respective standard deviations. Last line indicates the average considering all subsets. In training stage, different processing times were verified among the subsets since the number of images between them are different. Table 13 presents the average processing time for both stages and configurations considering FERET database.

For both datasets, it is possible to realize from the standard deviation very distinct execution times. This occurs because different techniques require different processing capabilities and thus, require more or less processing time. As to test a face image, a low standard deviation was verified since the images are already processed using the optimized set of techniques and parameters, this way taking milliseconds to classify an image. The main objective of employing DE$_B$ is to reduce training time. Based on this, we verified a drop of about 92% and 41% for CMU-PIE and FERET databases, respectively, without interfering significantly on test time.

Table 12 – The average processing time in training stage and to test a face image for each CMU-PIE subset - case study 2

	DE$_A$		**DE$_B$**	
	Training	**Test per face image**	**Training**	**Test per face image**
C05	11h 40m ± 9h 43m	0.0792s ± 0.0268s	0h 1m 43s ± 0h 0m 17s	0.0760s ± 0.0234s
C07	5h 2m ± 10h 16m	0.0117s ± 0.0279s	0h 34m 46s ± 0h 41m 15s	0.0416s ± 0.0399s
C09	7h 19m ± 3h 5m	0.0735s ± 0.0367s	0h 3m 20s ± 0h 3m 2s	0.0588s ± 0.0375s
C27	9h 9m ± 0h 10m	0.0895s ± 0.0010s	2h 23m 3s ± 1h 5m 42s	0.0571s ± 0.0775s
C29	9h 3m ± 0h 27m	0.0911s ± 0.0014s	0h 2m 14s ± 0h 0m 27s	0.0735s ± 0.0245s
C$_{All}$	46h 58m ± 16h 42m	0.0963s ± 0.0368s	8h 11m 46s ± 6h 16m 45s	0.1043s ± 0.0059s
Average	14h 51m ± 15h 52m	0.0735s ± 0.0314s	1h 52m 48s ± 1h 21m 14s	0.0587s ± 0.0347s

Source: Own authorship

Table 13 – The average processing time in training stage and to test a face image for each FERET subset - case study 2

	Training	**Test per face image**
DE$_A$	5h 59m 00s ± 1h 50m 00s	0.0075s ± 0.0009s
DE$_B$	3h 25m 15s ± 1h 17m 29s	0.0152s ± 0.0069s

Source: Own authorship

Regarding the techniques and parameters employed by the FR framework, Tables 14 and 15 present the settings chosen by the DE$_A$ and DE$_B$ algorithms in the optimization step for the C$_{All}$ and FERET databases, respectively. Each row represents the setting found in each experiment, in the same order as the recognition rates are presented. Each column represents a parameter of the employed techniques. A dash symbol means that the related technique was not chosen. Since similar conclusions is draw from the CMU-PIE subsets, we only present the settings for C$_{All}$ subset. For this subset, it is possible to realize that the WBIN technique was employed in every

Table 14 – Settings from training step on C$_{All}$ subset - case study 2

(a) DE$_A$

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	sym20	0.0011	3	-	-	-	-	-	-	-	-	-	Yes	L_1
2	db17	0.2592	-	-	-	-	-	-	-	-	-	-	Yes	L_1
3	db29	0.3012	-	-	-	-	-	-	-	-	-	-	Yes	L_1
4	db20	0.1155	2.7979	-	-	-	-	-	-	-	-	-	Yes	L_1
5	sym8	0.9618	2.6335	-	-	-	5	1.3975	1.4946	-	-	-	-	L_2
6	db27	0.4953	-	-	-	-	-	-	-	-	-	-	Yes	L_1
7	db25	0.1258	-	-	-	-	-	-	-	-	-	-	Yes	L_1
8	sym20	0.0023	-	-	-	-	-	-	-	-	-	-	Yes	L_1
9	sym10	0.9165	3	3	1	2.9235	5	1.6527	2.5990	coif17	1	Yes	-	L_{2sqr}
10	sym20	0	1.6181	-	-	-	-	-	-	-	-	Yes	Yes	L_1

(b) DE$_B$

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	-	-	3.0000	3	1.0223	1.8838	-	-	-	-	-	-	Yes	L_1
2	coif7	0.6189	1.6946	-	-	-	-	-	-	-	-	-	Yes	L_1
3	sym20	0.0000	-	-	-	-	-	-	-	-	-	Yes	Yes	L_2
4	db20	0.0000	-	-	-	-	-	-	-	-	-	Yes	Yes	L_1
5	db9	0.3188	-	-	-	-	-	-	-	-	-	-	Yes	L_1
6	db9	0.3188	-	-	-	-	-	-	-	-	-	-	Yes	L_1
7	sym20	0.8224	2.3678	-	-	-	-	-	-	-	-	-	Yes	L_1
8	sym16	0.0000	2.0656	-	-	-	-	-	-	-	-	Yes	Yes	L_2
9	db14	0.0000	-	-	-	-	-	-	-	-	-	-	Yes	L_1
10	-	-	3.0000	3	1.6786	1.2674	-	-	-	-	-	-	Yes	L_1

Source: Own authorship

experiment performed but one, considering both algorithm's configurations, and the Manhattan distance (L_1) was the preferred choice for the distance measure. Also, LBP was employed in almost every case, and when it was not, the Laplacian was employed to somehow compensate dataset conditions. Regarding FERET database, we can observe that for both cases the algorithm tends to choose the GIC and Gaussian Blur as preprocessing techniques, ignoring the Laplacian and WBIN. The one single case where the DE$_B$ chooses WBIN, could be considered as an outlier. For feature extraction, the preferred technique employed was the DWT. It worth to point out that in this case the Manhattan distance was chose in every experiment. Despite these patterns, distinct settings were chosen by both algorithms for CMU-PIE and FERET datasets considering the techniques and parameters.

The convergence and diversity charts for DE$_A$ and DE$_B$ on C$_{All}$ and FERET databases are presented in Figures 18 and 19, respectively. It is possible to observe that the algorithm does not improve substantially around the 40th iteration for both datasets. However, some subsets from CMU-PIE are rather trivial achieving 100% recognition rate in the first iteration for every experiment e.g. C05. In relation to diversity, the same behavior as in Case study 1 (Section 6.2) was verified, therefore presenting a high diversity during all run.

In this case study, we verified that the proposed framework is robust across

Table 15 – Settings from training step on FERET dataset - case study 2

(a) DE$_A$

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	-	-	-	31	5	1.0274	-	-	-	sym4	3	-	-	L_1
2	-	-	1.3620	-	-	-	-	-	-	rbio3.1	4	-	-	L_1
3	-	-	1.0452	15	5	3.6923	-	-	-	rbio2.2	4	-	-	L_1
4	-	-	1.1127	25	4.2483	1	-	-	-	sym4	3	-	-	L_1
5	-	-	1.2315	9	1.3918	1.1867	-	-	-	rbio3.1	4	-	-	L_1
6	-	-	1.4304	31	4.5834	3.2495	-	-	-	rbio3.3	3	-	-	L_1
7	-	-	-	19	3.9	1.7166	-	-	-	rbio2.2	4	-	-	L_1
8	-	-	-	25	4.8869	1.9915	-	-	-	sym4	3	-	-	L_1
9	-	-	1.2292	23	4.0069	4.3252	-	-	-	sym4	3	-	-	L_1
10	-	-	-	3	1.5225	2.2343	-	-	-	rbio3.1	4	-	-	L_1

(b) DE$_B$

Experiment	WBIN		GIC	Gaussian Blur			Laplacian			DWT		DCT	LBP	Distance Measure
	wf	sf	γ	$Kernel$	sdX	sdY	$Kernel$	sf	$Delta$	wf	$level$			
1	-	-	1.2645	27	4.6986	2.6978	-	-	-	sym4	3	-	-	L_1
2	-	-	-	21	5.0000	1.5034	-	-	-	bior3.7	2	Yes	-	L_1
3	sym17	0.0000	-	31	5.0000	1.0000	-	-	-	sym20	2	-	-	L_1
4	-	-	1.2424	23	4.7019	1.0000	-	-	-	coif6	2	-	-	L_1
5	-	-	3.0000	29	4.3895	5.0000	-	-	-	sym6	3	-	-	L_1
6	-	-	-	29	3.7766	3.0480	-	-	-	rbio2.6	3	-	-	L_1
7	-	-	1.5578	31	4.1918	2.4842	-	-	-	sym20	2	-	-	L_1
8	-	-	2.2776	-	-	-	-	-	-	rbio3.9	3	Yes	-	L_1
9	-	-	1.0460	13	3.7362	3.6576	-	-	-	-	-	-	-	L_1
10	-	-	-	31	4.9132	3.2616	-	-	-	rbio3.3	4	Yes	-	L_1

Source: Own authorship

different illumination conditions and pose variation, also it was proposed a different configuration for DE (DE$_B$) aiming to reduce training time. With the new configuration a trade-off was verified between accuracy and processing time. Following, in Case Study 3, we employed a self-adaptive version of DE called jDE with DE$_B$ configuration.

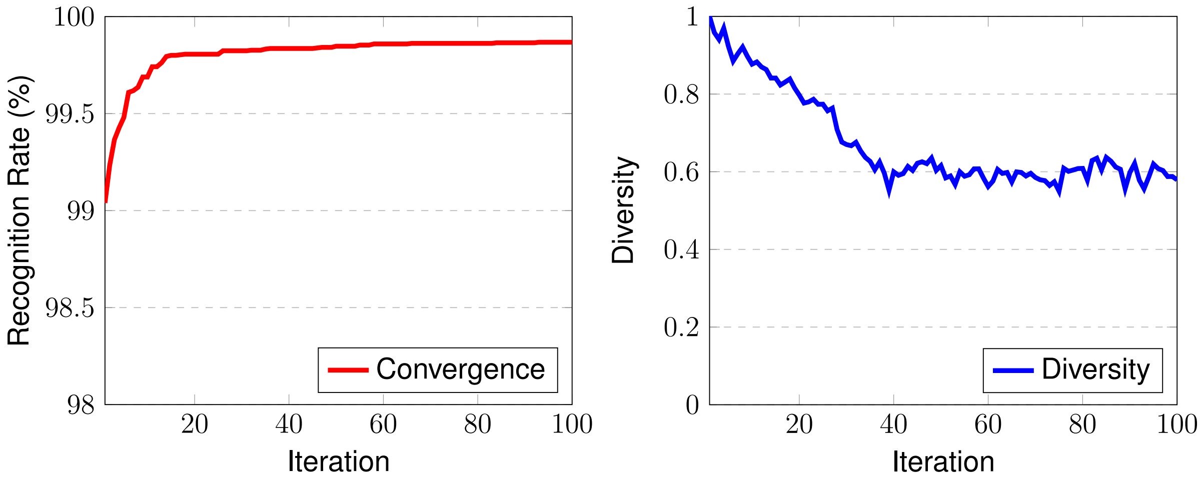
6.4 CASE STUDY 3

A self-adaptive version of DE, namely jDE, is employed in this case study. The main goal of this experiment is to get rid of the definition of the algorithm's parameters and lead the algorithms to a better search in the problem's space. Following, the results achieved are presented and compared with DE$_B$.

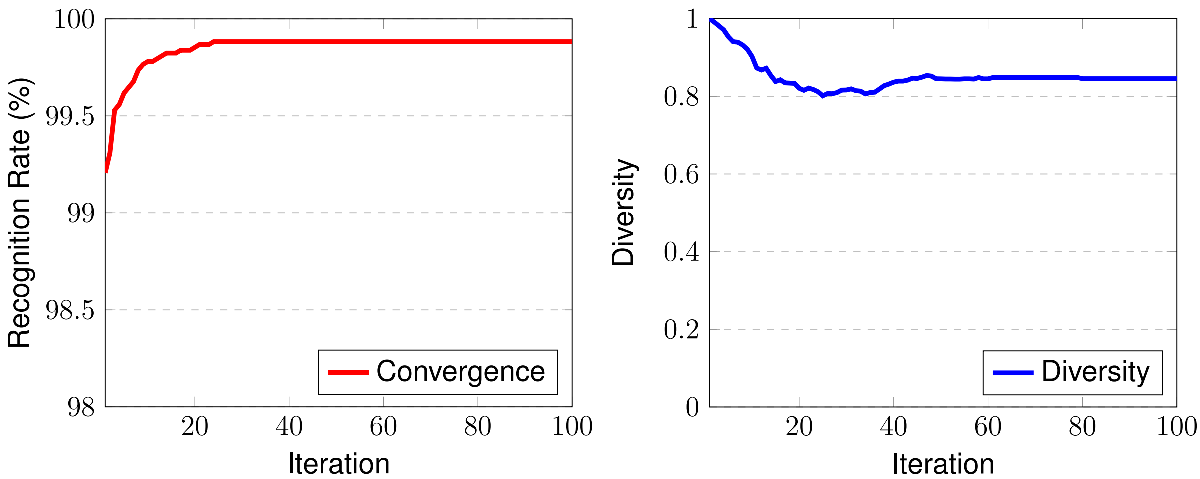
Table 16 presents the results obtained using jDE and DE$_B$ on each CMU-PIE subset and the average recognition rate. Regarding test stage, it is possible to realize that a gain in accuracy is achieved when jDE is employed for each subset except for C$_{All}$. Considering the statistical tests, only the 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 training stage, no statistical difference was verified. Table 17 shows the recognition rates for jDE and DE$_B$ considering FERET database. For this dataset, statistical difference in favor of jDE ($p - value = 0.0069$) was verified only in test stage.

Figure 18 – Convergence and Diversity charts, respectively, on c_{All} subset - case study 2

(a) DE_A



(b) DE_B



Source: Own authorship

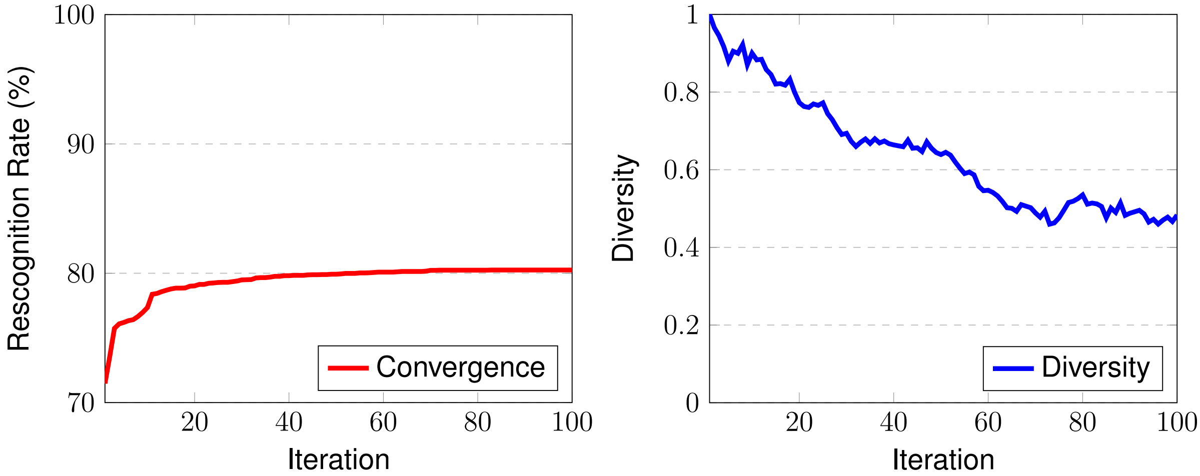
Table 16 – Recognition rates (%) using jDE and DE_B on CMU-PIE subsets - case study 3

	jDE		DE_B	
	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

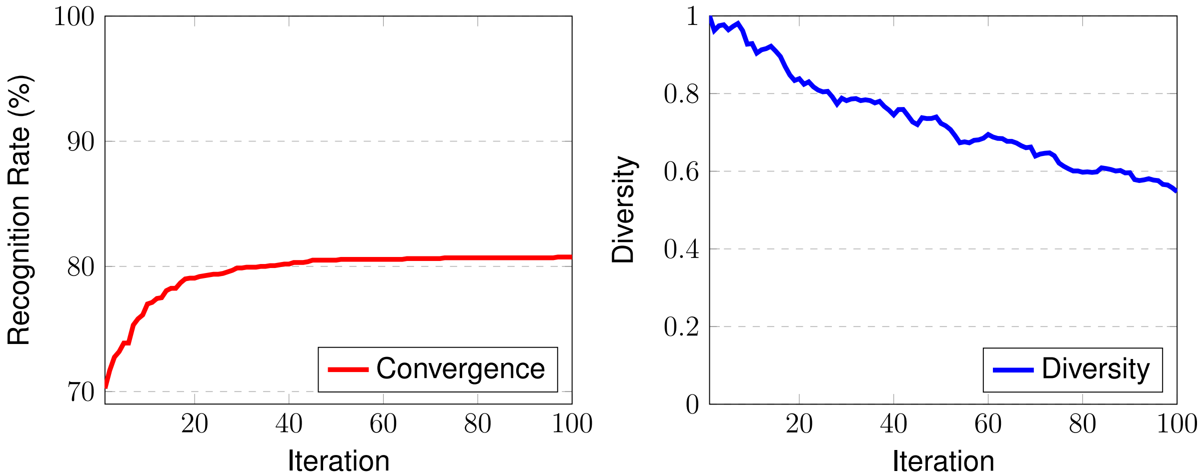
Source: Own authorship

Figure 19 – Convergence and Diversity charts, respectively, on FERET dataset - case study 2

(a) DE$_A$



(b) DE$_B$



Source: Own authorship

Regarding processing time, Tables 18 and 19 presents the average running times for each CMU-PIE subset and the FERET database, respectively. Considering the average processing time for CMU-PIE subsets, no significant difference was verified compared with DE$_B$. The same conclusions can be drawn for FERET database. Despite that, it is worth to point it out that an impressive gain in training is achieved when compared with DE$_A$ (Section 6.2).

Tables 20 and 21 show the settings found by jDE in training step for C$_{All}$ subset and FERET database, respectively. The experiments are in rows, while the techniques chosen are in columns. A dash symbol means that the technique was not employed

Table 17 – Recognition rates (%) using jDE and DE_B on FERET dataset - case study 3

	jDE		DE_B	
	Training	**Test**	**Training**	**Test**
FERET	80.56 ± 1.70	79.53 ± 0.22	80.75 ± 2.50	77.51 ± 2.26

Source: Own authoship

Table 18 – The average processing time in training and test stages on CMU-PIE subsets - case study 3

	jDE		DE_B	
	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
Average	1h 23m 20s ± 1h 5m 58s	0.0635 ± 0.0230	1h 52m 48s ± 1h 21m 14s	0.0587s ± 0.0347s

Source: Own authoship

Table 19 – The average processing time in training and test stages on FERET database - case study 3

	jDE		DE_B	
	Training	**Test per face image**	**Training**	**Test per face image**
FERET	2h 13m 37s ± 0h 42m 11s	0.0093s ± 0.0025s	3h 25m 15s ± 1h 17m 29s	0.0152s ± 0.0069s

Source: Own authoship

in the related experiment. Considering C_{All} subset, it is possible to realize that WBIN was chosen in every experiment except one, and the LBP technique was always chosen. Once more, 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), as we verified on case study 2. On the other side, the Gaussian Blur, Laplacian and DWT were never chosen. Different from CMU-PIE database, the Gaussian Blur and DWT was chosen in every experiment for FERET database. With this, we can verify our hyphotesis that different datasets require different techniques to deal with the its conditions. At last, we can observe that the Manhattan distance was once more the preferred distance measure.

Table 20 – Settings from training step C_{All} subset - case study 3

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}

Source: Own authoship

Figures 20 and 21 show the convergence and diversity charts for C_{All} subset

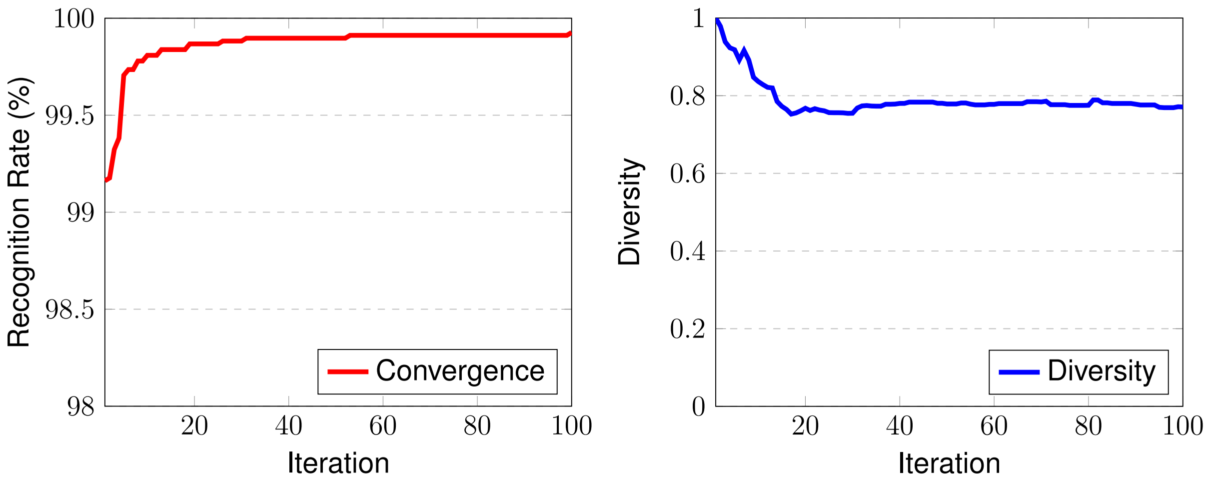
Table 21 – Settings from training step using on FERET dataset - case study 3

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

Source: Own authoship

and FERET database, respectively. Once more, only the charts of C_{All} subset are presented because the same conclusions can be drawn from the other subsets. As we can see, the same behavior as in the other experiments is verified. After around the 40^{th} iteration the solution does not improve significantly, and the diversity remains high during all algorithm's process.

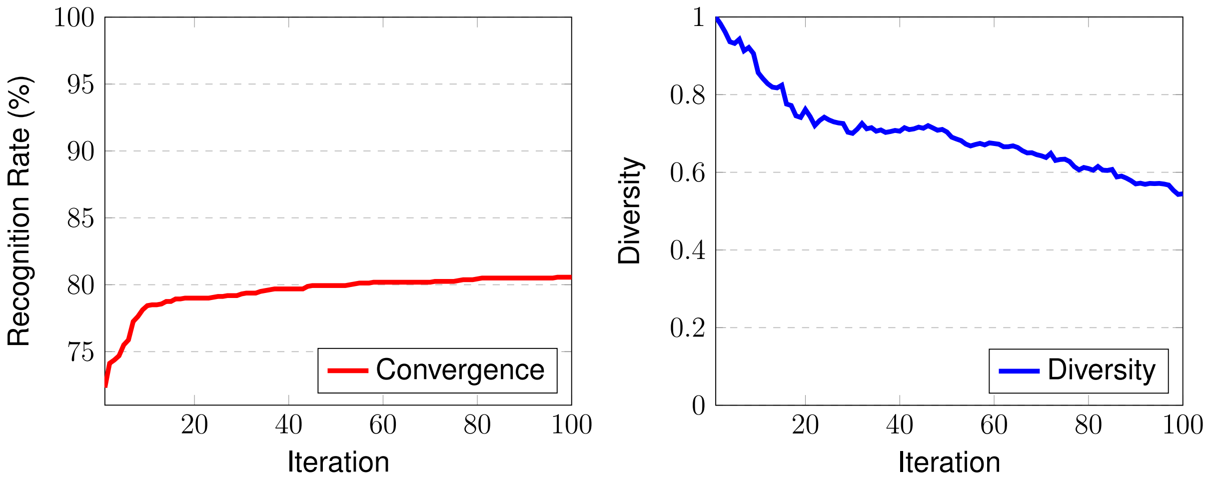
Figure 20 – Convergence and Diversity charts, respectively, on C_{All} - case study 3



Source: Own authoship

From this case study, it was possible to conclude that jDE achieved superior recognition rates when compared to DE_B, however the configuration employed on DE_A still achieved the highest recognition rates. It is worth to mention that experiments employing jDE with DE_A configuration were not performed due time availability. Considering processing time, the reduction of the validation images used on DE_B and jDE reflected in a considerably reduction on execution time as expected. Following, a comparison with the related work found in literature is presented.

Figure 21 – Convergence and Diversity charts, respectively, on FERET - case study 3



Source: Own authoship

6.5 COMPARISON WITH WORKS FOUND IN LITERATURE

In this subsection, the best results achieved by the proposed approach in the three databases, i.e. Yale Extended B, CMU-PIE and FERET, are compared with the results obtained by related works found in literature (Section 3.3). The best average result obtained by the proposed approach in training and test stages was chosen for comparison.

Table 22 presents the recognition rates obtained on Yale Extended B database. Among all experiments, the experiment 8 using DE optimization algorithm stood out with 100% of overall recognition rate on test stage. Table 22 presents the recognition rates obtained for each subset (S2-S5) and its average recognition rate (S$_{All}$). Also, 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. They are: for the wavelet family the haar wavelet is the most common; in DWT, only one level of decomposition was employed; the scale factor (SF) was set to 3 based on Manikantan et al. (2012) work; the parameter γ was set to 2; for Gaussian Blur, kernel size was set to 3x3 and the standard deviation based on kernel size; for Laplacian, the kernel size was also 3x3, and no scale or delta value were employed.

From Table 22, most of the results obtained from the literature presented a precision of 100% in subsets 2 and 3, which have better illumination. Consequently, this condition facilitates the illumination compensation. Meanwhile, the compensation on the subsets 4 and 5 are not that trivial since they contain the darkest images where the face is barely visible. However, most of the works have achieved recognition rates

Table 22 – Yale Extended B database recognition rates (%) of proposed approach compared to related works.

Method	S2	S3	S4	S5	S_{All}
DTCWT (BARADARANI; WU; AHMADI, 2013)	100	100	98.68	99.03	99.42
resonance (BARADARANI; WU, 2013)	100	100	98.68	99.16	99.46
DWT (HU, 2011)	100	100	98.86	99.35	99.55
WT-Retinal (CHENG; LI; JIAO, 2013)	99.78	99.44	98.47	97.78	98.86
fuzzy + LFDCT (VISHWAKARMA, 2015)	-	100	98.87	98.06	-
weighting GAP (ZHAO et al., 2013)	100	100	99.51	100	99.82
NMR (YANG et al., 2017)	-	-	90.2	47.9	-
CZT and Goertzel (VARADARAJAN et al., 2015)	-	-	-	96.25	-
Our method	**100**	**100**	**100**	**99.83**	**99.95**
Literature settings	84.20	68.65	21.48	48.02	55.59

Source: Own authorship

higher than 98% using both subsets. Among these, we can highlight the weighting GAP (sixth approach in Table 22) and our method achieving 99.95% recognition rate. From the results obtained from our approach employing literature settings, we can realize the importance of the parameter tuning within the optimized techniques and how efficient was the DE in optimizing the parameters.

Table 23 shows the recognition rates achieved by the proposed approach and related works on CMU-PIE database. The last line shows the results obtained when using the literature settings. It is worth to remember that all these works consider the OSPP problem, this way raising the difficulty of FR. 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 23, regarding the results achieved using literature settings, as different techniques reached the same highest recognition rates, the criterion used to selected the experiment to try out with literature settings was less processing time. For the C05 subset, experiment 5 using DE_A was chosen where only Laplacian technique was employed with Manhattan distance. So, we set the kernel size to 3x3 with no scale or delta value, and the Euclidean distance was employed. 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 1 using C07 dataset showed to be sensitive dropping 7.82% recognition rate when using literature settings. The γ parameter was set to 2, and for Laplacian parameters we set the same values used in subset C05 with Euclidean distance. Another example is experiment 5 from subset C27, which dropped slightly the performance ($\pm$1%) when literature settings were employed. In this, Gaussian Blur was employed besides GIC, DCT and LBP. For Gaussian Blur, the kernel size was set to 3x3 and the standard deviations were computed from kernel size. The actual impact of tuning parameters is shown in the remaining subsets i.e. C09, C29 and C$_{All}$, where a drop of about 25% was verified. In experiment 4 from C09, the techniques chosen by DE$_A$ algorithm were WBIN, Laplacian and LBP with Manhattan distance. For WBIN, the haar wavelet with scale factor 3 were considered the literature parameters. Experiment 10 using DE$_A$ was chosen to compare with literature settings using C29 subset. Also, when using literature settings in experiment for C$_{All}$ subset, a rate drop of 21.11% is verified.

Table 23 – CMU-PIE database recognition rates (%) of proposed approach compared to related works.

Method	C05	C07	C09	C27	C29	C$_{All}$
CD-EJSR (SHANG et al., 2018)	50	-	-	-	-	-
LRSR (GU; HU; LI, 2018)	-	93.46	93.63	-	-	-
R&L_ESRC (HU et al., 2017)	-	-	-	≈92.00	-	-
TLFL (GUO et al., 2017)	-	54.87	56.47	-	49.71	-
Our method	**100**	**99.92**	**100**	**99.77**	**100**	**99.90**
Literature settings	100	91.95	73.67	98.67	75.07	78.75

Source: Own authorship

Table 24 presents the results obtained by the related work and by the proposed FR framework on FERET database. Partial results of each subset and the average (*b*$_{All}$) are shown in columns, while proposed approaches are shown in rows. The last row shows results achieved using the techniques found by the framework but with parameter values commonly used literature. For comparison, the setup obtained by experiment 1 using DE$_A$ 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 (OH et al., 2018) 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. Moreover, once more its possible to draw from the results using literature settings the importance in tuning the parameters of the employed techniques.

Table 24 – FERET database recognition rates (%) of proposed approach compared to related works.

Method	*bb*	*bc*	*bd*	*be*	*bf*	*bg*	*bh*	*bi*	b_{All}
frequency SLPMM (MEHRASA; ALI; HOMAYUN, 2017)	15.5	19.5	43.5	77	82	51.5	28.5	14.5	41.5
AGFN$_{hor}$ (OH et al., 2018)	45.4	60.2	**93.6**	**97.6**	**99.3**	**96.7**	**78.2**	52.4	77.92
MFSA+ (CHU; ZHAO; AHMAD, 2018)	-	-	61	76.5	85	52	-	-	-
Our method	**68**	**82.5**	90	92	89	86.5	76	**60**	**80.5**
Literature settings	38	60.5	70.5	79	78.5	66.5	52.5	33.5	59.87

*Best results are highlighted in bold.

Source: Own authoship

7 CONCLUSIONS

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, in this Master thesis, we proposed a FR framework with the aid of bio-inspired optimization algorithms. The proposed approach implements several techniques for preprocessing and for feature extraction. The main aspect of this approach is the definition of which set of techniques and its ideal parameters values should be used for the FR using the bio-inspired optimization algorithm. In this work, we implemented several illumination compensation techniques to handle the problem of illumination variation, as it is a critical factor. Based on the literature review, the most used optimization algorithms are the PSO and the DE. Hence, we employ them for selecting the techniques, as well as to tune their parameters. The performance of the canonical versions of both algorithms was evaluated, and a self-adaptive version of DE was employed. The well known Yale Extended B, 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 the results obtained from case study 1, it can be said that in training stage the DE is statistically better then PSO when considering accuracy. On the other hand, no statistical differences were verified in test stage. Regarding processing time, no significant differences were verified. Despite that, the box-plots made possible to realize the superiority of DE algorithm with more consistent results for both time and accuracy. This way, our approach demonstrated suitability for compensating illumination variation using Yale Extended B face dataset and it presents competitive results in comparison with the works proposed in the literature. The subsets 2 and 3 of the dataset have better illumination compared to 4 and 5, therefore it is trivial to compensate them. On the other hand, the challenge remains on the subsets 4 and 5 in which we can note the presence of the dark and partially illuminated images. According to the results obtained from the experiments using these subsets, our approach reached almost 100% of recognition rate. It strongly demonstrates the robustness of our method. Hence, both proposed algorithms has shown to be suitable for this kind of task.

Since it was verified that DE algorithm is superior than PSO for this problem, only DE was employed in case study 2. However, a new configuration was proposed (DE$_B$). In this, we reduced the number of the validation images to 10% and defined 100% recognition rate as stop criterion to reduce training time, after, we compared with the traditional configuration (DE$_A$). Regarding the experiments performed on CMU-PIE subsets using DE$_A$, the proposed approach showed to be suitable in compensation illumination variation achieving at least 99% of recognition rate. 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 slightly changes of pose as well. Comparing with DE$_B$, a loss in accuracy was verified. For subset C27, a recognition drop of about 3% and 4% is verified for training and test stages, respectively. In addition, a lost in precision is observed, specially on subsets C09 and C29. Regarding training time, a drop of about 92% was verified considering the average among the subsets. It is worth to point out that the OSPP problem was considered for both datasets. For FERET database using DE$_A$, considerably recognition rates were found, since only techniques that deals with illumination were implemented (more or less 80%). However, with DE$_B$, an accuracy loss of about 4% was verified. On the other hand, 41% of gain in training time was achieved. With this, it becomes clear a trade-off between accuracy and processing time considering both configurations. In comparison with literature, our approach demonstrated competitive results.

In case study 3, we employed a self-adaptive version of DE, namely jDE, with DE$_B$ configuration. In this, we attempt to compensate the loss of accuracy verified using DE$_B$. The jDE achieved superior recognition rates when compared to DE$_B$, however the configuration employed on DE$_A$ still achieved the highest recognition rates.

It was possible to realize in experiments that optimizing the set of techniques for illumination compensation and for feature extraction as well as their parameters is important to enhance the precision in a FR system. This is evident when taking a look at the experiment using literature settings in which the use of the same set of techniques with optimized parameter values achieved much higher recognition rate. Also, through the experiments, it can be observed that the present FR problem is highly multi-modal since different sets of techniques can achieve similar (or the same) results. From the aimed objectives to accomplished in this work (Section 1.2), it can be said that both optimization algorithms are suitable to the task of finding an optimized solution and that the implemented techniques presented the required strength for compensating illumination variation on the employed dataset. Following, we point out some future directions to handle other problems found in FR process and enhance the proposed system.

7.1 FUTURE WORKS

To complement this work, we found that some future implementations are necessary to develop a robust and complete FR system. Following, we list these directions:

- As different strategies can give approximate the same recognition rates but with different processing times, we suggest to use the processing time and recognition rate as optimization criteria.

- In this work, we only consider a specified sequence of the employed preprocessing techniques. We suggest to include the possibility of interchange these techniques in the optimization step.
- 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 added to the framework and other datasets should be experimented.
- The use of High Performance Computing (HPC). 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 massive parallel hardware architectures, such as GPUs, could be used.

7.2 CONTRIBUTIONS

During the research period of this project, it was developed and published some works that built the steps towards the construction and accomplishment of this master thesis.

The first work aimed to compensate illumination variation using a homomorphic filter with optimized parameters provided by PSO and Jaya (RAO, 2016) algorithms (PLICHOSKI; PARPINELLI; CHIDAMBARAM, 2017). The main objective of this work was to evaluate the performance of both algorithms in finding an optimized solution, as well as to promote the strength of the homomorphic filter.

In the next work, we investigate unlighting photometric-based techniques together with the LBP texture descriptor for illumination compensation (PLICHOSKI; PARPINELLI; CHIDAMBARAM, 2018b). A cascade of preprocessing and feature extraction techniques was employed achieving the highest recognition rate on Yale Extended B database considering the experimental methodology employed.

Next, a study of bio-inspired approaches for optimizing FR systems became necessary. With this, we performed an extensive Systematic Literature Review (SLR) contemplating works that employed bio-inspired algorithms from EC and SI in 2D-FR systems in the range of 2012-2018 (PLICHOSKI; PARPINELLI; CHIDAMBARAM, 2018c). We focused in what and how the optimization procedure was employed in each reviewed work, thus gathering seventy three papers with a variety of algorithms for different purposes.

Based on the previous systematic literature review, we presented an adjustable FR framework with the aid of the DE (PLICHOSKI; PARPINELLI; CHIDAMBARAM, 2018a). This framework contains all the preprocessing and feature extraction techniques proposed in Plichoski, Parpinelli and Chidambaram (2018b) and DE is responsible for choosing which strategies to use, as well as tuning the parameters involved.

After, the comparison between the two most used bio-inspired algorithms i.e. PSO and DE, accordingly to our systematic literature review, were performed in the proposed framework (PLICHOSKI; CHIDAMBARAM; PARPINELLI, 2018a). Basically, it comprehends case study 1.

An extensive review about on-line parameter control strategies applied on SI and EC-based bio-inspired algorithms was performed (PARPINELLI et al., 2018). We verified that the setup of parameters has an important role in defining their behavior, guiding the search and biasing the quality of final solutions. With this, we intend to apply an on-line control for DE parameters in the optimization procedure of the FR system proposed in this work as stated in Section 7.1.

Besides these works, another work was sent to peer-review in the Swarm and Evolutionary Computation journal. This work contemplates most of this master thesis, except the experiments employing jDE (PLICHOSKI; CHIDAMBARAM; PARPINELLI, 2018b).

BIBLIOGRAPHY

ABATE, A. F. et al. 2D and 3D face recognition: A survey. **Pattern recognition letters**, Elsevier, v. 28, n. 14, p. 1885–1906, 2007.

ABHISHREE, T. et al. Face recognition using gabor filter based feature extraction with anisotropic diffusion as a pre-processing technique. **Procedia Computer Science**, Elsevier, v. 45, p. 312–321, 2015.

ADINI, Y.; MOSES, Y.; ULLMAN, S. Face recognition: The problem of compensating for changes in illumination direction. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 19, n. 7, p. 721–732, 1997.

AJAYA, H. S. et al. Face recognition by optimal band selection in contourlet transform domain. In: ABV-IIITM. **2014 9th International Conference on Industrial and Information Systems (ICIIS)**. Gwalior: IEEE, 2014. p. 1–8.

ALSALIBI, B. et al. The impact of bio-inspired approaches toward the advancement of face recognition. **ACM Computing Surveys (CSUR)**, ACM, v. 48, n. 1, p. 5, 2015.

ANDRÉ, L.; PARPINELLI, R. S. Tutorial sobre o uso de técnicas para controle de parâmetros em algoritmos de inteligência de enxame e computação evolutiva. **Revista de Informática Teórica e Aplicada**, v. 21, n. 2, p. 90–135, 2014.

ANEESH, M.; MASAND, A. A.; MANIKANTAN, K. Optimal feature selection based on image pre-processing using accelerated binary particle swarm optimization for enhanced face recognition. **Procedia Engineering**, Elsevier, v. 30, p. 750–758, 2012.

BABU, S. H.; BIRAJDHAR, S. A.; TAMBAD, S. Face recognition using entropy based face segregation as a pre-processing technique and conservative BPSO based feature selection. In: **Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing**. New York, NY, USA: ACM, 2014. (ICVGIP '14), p. 46:1–46:8.

BABU, S. H. et al. Face recognition using active illumination equalization and mirror image superposition as pre-processing techniques. In: B.N.M INSTITUTE OF TECHNOLOGY. **2014 Fifth International Conference on Signal and Image Processing**. Bangalore: CPS, 2014. p. 96–101.

BANERJEE, P. K.; DATTA, A. K. A preferential digital–optical correlator optimized by particle swarm technique for multi-class face recognition. **Optics & Laser Technology**, Elsevier, v. 50, p. 33–42, 2013.

BARADARANI, A.; WU, Q. J.; AHMADI, M. An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering. **Pattern Recognition**, Elsevier, v. 46, n. 1, p. 57–72, 2013.

BARADARANI, A.; WU, Q. M. J. Illumination invariant human face recognition: frequency or resonance? In: IEEE. **2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. Shanghai: IEEE, 2013. p. 1–6.

BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 19, n. 7, p. 711–720, 1997.

BENITEZ, C. M. V.; PARPINELLI, R. S.; LOPES, H. S. Parallelism, hybridism and coevolution in a multi-level ABC-GA approach for the protein structure prediction problem. **Concurrency and Computation: Practice and Experience**, Wiley Online Library, v. 24, n. 6, p. 635–646, 2012.

BOWYER, K. W.; CHANG, K.; FLYNN, P. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. **Computer vision and image understanding**, Elsevier, v. 101, n. 1, p. 1–15, 2006.

BREST, J.; ZUMER, V.; MAUCEC, M. S. Self-adaptive differential evolution algorithm in constrained real-parameter optimization. In: 555. **Evolutionary Computation, 2006. CEC 2006. IEEE Congress on**. Vancouver, BC, Canada, 2006. p. 215–222.

BROWNLEE, J. **Clever Algorithms: Nature-Inspired Programming Recipes**. 1st. ed. Melbourne: Lulu.com, 2011. ISBN 1446785068, 9781446785065.

CAPES. Documento de área 2013. **Brasília, DF: Capes**, 2013.

CHEN, L.; MAN, H.; NEFIAN, A. V. Face recognition based on multi-class mapping of fisher scores. **Pattern Recognition**, Elsevier, v. 38, n. 6, p. 799–811, 2005.

CHENG, Y. et al. Illumination invariant face recognition with particle swarm optimization. In: IEEE. **Data Mining Workshop (ICDMW), 2014 IEEE International Conference on**. Shenzhen: CPS, 2014. p. 862–866.

CHENG, Y.; LI, Z.; JIAO, L. Illumination normalization for face recognition under extreme lighting conditions. In: YANG, J.; FANG, F.; SUN, C. (Ed.). **Intelligent Science and Intelligent Data Engineering**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 491–497. ISBN 978-3-642-36669-7.

CHIDAMBARAM, C. et al. Multiple face recognition using local features and swarm intelligence. **IEICE TRANSACTIONS on Information and Systems**, The Institute of Electronics, Information and Communication Engineers, v. 97, n. 6, p. 1614–1623, 2014.

CHU, Y.; ZHAO, L.; AHMAD, T. Multiple feature subspaces analysis for single sample per person face recognition. **The Visual Computer**, Springer, p. 1–18, 2018.

CORRIVEAU, G. et al. Review of phenotypic diversity formulations for diagnostic tool. **Applied Soft Computing**, Elsevier, v. 13, n. 1, p. 9–26, 2013.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE transactions on information theory**, IEEE, v. 13, n. 1, p. 21–27, 1967.

DARESTANI, M. R. Y.; SHEIKHAN, M.; KHADEMI, M. Face recognition using contourlet-based features and hybrid PSO-neural model. In: **The 5th Conference on Information and Knowledge Technology**. Shiraz, Iran: IEEE, 2013. p. 1818–186.

DAS, S.; ABRAHAM, A.; KONAR, A. Swarm intelligence algorithms in bioinformatics. **Computational Intelligence in Bioinformatics**, Springer, p. 113–147, 2008.

DASGUPTA, D.; YU, S.; NINO, F. Recent advances in artificial immune systems: models and applications. **Applied Soft Computing**, Elsevier, v. 11, n. 2, p. 1574–1587, 2011.

D'CUNHA, N. W. et al. Face recognition using homomorphic filtering as a pre-processing technique. In: **Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on**. Bangalore, India: IEEE, 2013. p. 1–6.

DEEPA, G. et al. Face recognition using spectrum-based feature extraction. **Applied Soft Computing**, Elsevier, v. 12, n. 9, p. 2913–2923, 2012.

DETROZ, J. P.; HINZ, M.; HOUNSELL, M. S. The use of literature review in informatics in education: a systematic mapping. **Brazilian Journal of Computers in Education**, v. 23, n. 01, p. 28, 2015.

DIVYA, R. et al. Astroid shaped DCT feature extraction for enhanced face recognition. In: **Proceedings of the CUBE International Information Technology Conference**. New York, NY, USA: ACM, 2012. p. 95–101.

DOGRA, A. Performance comparison of different wavelet families based on bone vessel fusion. **Asian Journal of Pharmaceutics (AJP)**, v. 10, n. 04, 2017.

DORIGO, M.; STUTZLE, T. The ant colony optimization metaheuristic: Algorithms, applications, and advances. **International series in operations research and management science**, Springer, p. 251–286, 2003.

DU, S.; WARD, R. Wavelet-based illumination normalization for face recognition. In: **Image Processing, 2005. ICIP 2005. IEEE International Conference on**. Genova, Italy: IEEE, 2005. p. II–954.

EBERHART, R. C.; SHI, Y. **Computational intelligence: concepts to implementations**. Burlington: Morgan Kaufmann, 2011. v. 1.

EROL, O. K.; EKSIN, I. A new optimization method: Big bang-big crunch. **Advances in Engineering Software**, v. 37, n. 2, p. 106 – 111, 2006.

FAN, X.; VERMA, B. Face recognition: a new feature selection and classification technique. In: **Proc. 7th Asia-Pacific Conference on Complex Systems**. Cairns, Australia: Rockhampton, 2004. p. 713–721.

FARAG, M. M. M.; ELGHAZALY, T.; HEFNY, H. A. Face recognition system using HMM-PSO for feature selection. In: **Computer Engineering Conference (ICENCO), 2016 12th International**. Cairo, Egypt: IEEE, 2016. p. 105–110.

FERNÁNDEZ-MARTÍNEZ, J. L.; CERNEA, A. Exploring the uncertainty space of ensemble classifiers in face recognition. **International Journal of Pattern Recognition and Artificial Intelligence**, World Scientific, v. 29, n. 03, p. 1556002, 2015.

FISTER, I.; BREST, J.; MLAKAR, U. Towards the universal framework of stochastic nature-inspired population-based algorithms. In: **Computational Intelligence (SSCI), 2016 IEEE Symposium Series on**. Athens, Greece: IEEE, 2016. p. 1–8.

Fister JR., I. et al. A Brief Review of Nature-Inspired Algorithms for Optimization. **ArXiv e-prints**, jul. 2013.

GAGAN, R. et al. Circular sector dct based feature extraction for enhanced face recognition using histogram based dynamic gamma intensity correction. In: **Proceedings of the CUBE International Information Technology Conference**. New York, NY, USA: ACM, 2012. p. 74–81.

GAO, W. et al. The CAS-PEAL large-scale chinese face database and baseline evaluations. **IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans**, IEEE, v. 38, n. 1, p. 149–161, 2008.

GEORGHIADES, A. S.; BELHUMEUR, P. N.; KRIEGMAN, D. J. From few to many: Illumination cone models for face recognition under variable lighting and pose. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 23, n. 6, p. 643–660, 2001.

GONZALEZ, R. C.; WOODS, R. E. Digital image processing, pretice hall, 2008. **ISBN-13**, 2008.

GOURIER, N.; HALL, D.; CROWLEY, J. L. Estimating face orientation from robust detection of salient facial structures. **FG Net Workshop on Visual Observation of Deictic Gestures**, Cambridge, UK, 2004.

GRAPS, A. An introduction to wavelets. **IEEE computational science and engineering**, IEEE, v. 2, n. 2, p. 50–61, 1995.

GROSS, R. et al. Multi-pie. **Image and Vision Computing**, Elsevier, v. 28, n. 5, p. 807–813, 2010.

GU, J.; HU, H.; LI, H. Local robust sparse representation for face recognition with single sample per person. **IEEE/CAA Journal of Automatica Sinica**, IEEE, v. 5, n. 2, p. 547–554, 2018.

GUI, J. et al. Feature selection based on structured sparsity: A comprehensive study. **IEEE transactions on neural networks and learning systems**, IEEE, v. 28, n. 7, p. 1490–1507, 2017.

GUO, Y. et al. Fuzzy sparse autoencoder framework for single image per person face recognition. **IEEE transactions on cybernetics**, IEEE, 2017.

HASAN, M.; ABDULLAH, S. N. H. S.; OTHMAN, Z. A. Face recognition based on opposition particle swarm optimization and support vector machine. In: **Signal and Image Processing Applications (ICSIPA), 2013 IEEE International Conference on**. Melaka, Malaysia: IEEE, 2013. p. 417–424.

HATAMLOU, A. Black hole: A new heuristic optimization approach for data clustering. **Information Sciences**, v. 222, n. 0, p. 175 – 184, 2013.

HEGADE, P. P. et al. DWT-based face recognition using morphological edge detection as a pre-processing technique. In: **Engineering (NUiCONE), 2013 Nirma University International Conference on**. Ahmedabad, India: IEEE, 2013. p. 1–6.

HOLLAND, J. H. Genetic algorithms and the optimal allocation of trials. **SIAM Journal on Computing**, SIAM, v. 2, n. 2, p. 88–105, 1973.

HU, C. et al. Illumination robust single sample face recognition based on ESRC. **Multimedia Tools and Applications**, Springer, v. 76, n. 24, p. 26523–26550, 2017.

HU, H. Variable lighting face recognition using discrete wavelet transform. **Pattern Recognition Letters**, Elsevier, v. 32, n. 13, p. 1526–1534, 2011.

HUANG, G. B. et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In: **Workshop on faces in'Real-Life'Images: detection, alignment, and recognition**. Marseille, France: HAL-Inria, 2008.

ISLAM, S. et al. A review of recent advances in 3D ear-and expression-invariant face biometrics. **ACM Computing Surveys (CSUR)**, ACM, v. 44, n. 3, p. 14, 2012.

ITSEEZ. **The OpenCV Reference Manual**. 3.4.1. ed. 11, 2018. Disponível em: <https://docs.opencv.org/>.

JAIN, V. The indian face database/vidit jain, amitabha mukherjee. **Kanpur: Indian Institute of Technology Kanpur**, 2002.

JESORSKY, O.; KIRCHBERG, K. J.; FRISCHHOLZ, R. W. Robust face detection using the hausdorff distance. In: BIGUN, J.; SMERALDI, F. (Ed.). **Audio- and Video-Based Biometric Person Authentication**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 90–95.

JONG, K. D. **Evolutionary Computation: A Unified Approach**. London, England: Mit Press, 2006. v. 1.

KANAN, H. R.; FAEZ, K.; HOSSEINZADEH, M. Face recognition system using ant colony optimization-based selected features. In: **Computational Intelligence in Security and Defense Applications, 2007. CISDA 2007. IEEE Symposium on**. Honolulu, HI, USA: IEEE, 2007. p. 57–62.

KARABOGA, D. An idea based on honey bee swarm for numerical optimization. Technical report, 2005.

KASINSKI, A.; FLOREK, A.; SCHMIDT, A. The PUT face database. **Image Processing and Communications**, v. 13, n. 3-4, p. 59–64, 2008.

KAUR, H.; PANCHAL, V.; KUMAR, R. A novel approach based on nature inspired intelligence for face feature extraction and recognition. In: **Contemporary Computing (IC3), 2013 Sixth International Conference on**. Noida, India: IEEE, 2013. p. 149–153.

KHADHRAOUI, T. et al. Features selection based on modified PSO algorithm for 2D face recognition. In: **Computer Graphics, Imaging and Visualization (CGiV), 2016 13th International Conference on**. Beni Mellal, Morocco: IEEE, 2016. p. 99–104.

KHAN, N.; GUPTA, M. Face recognition system using improved artificial bee colony algorithm. In: **Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on**. Chennai, India: IEEE, 2016. p. 3731–3735.

KHASHU, S. et al. Face recognition using dual wavelet transform and filter-transformed flipping. In: **Electronics and Communication Systems (ICECS), 2014 International Conference on**. Coimbatore, India: IEEE, 2014. p. 1–7.

KIM, S.-H.; OH, S.-K.; KIM, J.-Y. Design of face recognition system realized with the aid of PCA-based RBFNN. In: **Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems, 2016 Joint 8th International Conference on**. Sapporo, Japan: IEEE, 2016. p. 754–758.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. **Science**, v. 220, n. 4598, p. 671–680, 1983.

KISHORE, B. et al. Face recognition using gabor-feature-based DFT shifting. In: **Industrial and Information Systems (ICIIS), 2014 9th International Conference on**. Gwalior, India: IEEE, 2014. p. 1–8.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. 2007.

KODANDARAM, R. et al. Face recognition using truncated transform domain feature extraction. **International Arab Journal of Information Technology (IAJIT)**, v. 12, n. 3, 2015.

KONG, S. G. et al. Recent advances in visual and infrared face recognition—a review. **Computer Vision and Image Understanding**, Elsevier, v. 97, n. 1, p. 103–135, 2005.

KRISSHNA, N. A. et al. Face recognition using transform domain feature extraction and PSO-based feature selection. **Applied Soft Computing**, Elsevier, v. 22, p. 141–161, 2014.

KUSHWAH, A. et al. A review: Comparative study of edge detection techniques. **International Journal**, v. 8, n. 5, 2017.

LEI, F. et al. Fast static particle swarm optimization based feature selection for face detection. In: **Computational Intelligence and Security (CIS), 2012 Eighth International Conference on**. Guangzhou, China: IEEE, 2012. p. 401–405.

LILJA, D. J. **Measuring computer performance: a practitioner's guide**. Cambridge, Reino Unido: Cambridge university press, 2005. v. 1.

LIU, L. et al. Local binary features for texture classification: Taxonomy and experimental study. **Pattern Recognition**, Elsevier, v. 62, p. 135–160, 2017.

LIU, N.; WANG, H. Feature extraction with genetic algorithms based nonlinear principal component analysis for face recognition. In: **Pattern Recognition, 2006. ICPR 2006. 18th International Conference on**. Hong Kong, China: IEEE, 2006. v. 3, p. 461–464.

LIU, X. et al. Study on color space for the performance of degraded face image recognition. In: **Media Watermarking, Security, and Forensics Conference, IS&T Electonic Imaging 2018**. Villeurbanne, France: HAL, 2018.

LIU, X. et al. A novel face recognition algorithm via weighted kernel sparse representation. **Future Generation Computer Systems**, Elsevier, v. 80, p. 653–663, 2018.

LYONS, M. et al. Coding facial expressions with gabor wavelets. In: **Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on**. Nara, Japan: IEEE, 1998. p. 200–205.

MALLIPEDDI, R.; LEE, M. Ensemble based face recognition using discriminant PCA features. In: **Evolutionary Computation (CEC), 2012 IEEE Congress on**. Brisbane, QLD, Australia: IEEE, 2012. p. 1–7.

MANIKANTAN, K. et al. DWT-based illumination normalization and feature extraction for enhanced face recognition. **International Journal of Engineering & Technology**, v. 1, n. 4, p. 483–504, 2012.

MARTINEZ, A. M. The AR face database. **CVC technical report**, 1998.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115–133, 1943.

MEHRASA, N.; ALI, A.; HOMAYUN, M. A supervised multimanifold method with locality preserving for face recognition using single sample per person. **Journal of Central South University**, Springer, v. 24, n. 12, p. 2853–2861, 2017.

MOLLAEE, M.; MOATTAR, M. H. Face recognition based on modified discriminant independent component analysis. In: **Computer and Knowledge Engineering (ICCKE), 2016 6th International Conference on**. Mashhad, Iran: IEEE, 2016. p. 60–65.

MURTHY, N. N. et al. Face recognition using DWT thresholding based feature extraction with laplacian-gradient masking as a pre-processing technique. In: **Proceedings of the CUBE International Information Technology Conference**. New York, NY, USA: ACM, 2012. p. 82–89.

NEBTI, S.; BOUKERRAM, A. Swarm intelligence inspired classifiers for facial recognition. **Swarm and Evolutionary Computation**, Elsevier, v. 32, p. 150–166, 2017.

NEMA, S.; THAKUR, S. Improved particle swarm optimization approach for classification by using LDA. In: **Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on**. Coimbatore, India: IEEE, 2015. p. 1–5.

NISCHAL, K. et al. Face recognition using entropy-augmented face isolation and image folding as pre-processing techniques. In: **India Conference (INDICON), 2013 Annual IEEE**. Mumbai, India: IEEE, 2013. p. 1–6.

NIVEDITHA, G. et al. DWT-based face recognition using fast walsh hadamard transform and chiral image superimposition as pre-processing techniques. In: **Electronics and Communication Systems (ICECS), 2015 2nd International Conference on**. Coimbatore, India: IEEE, 2015. p. 7–14.

OCHOA-VILLEGAS, M. A. et al. Addressing the illumination challenge in two-dimensional face recognition: a survey. **IET Computer Vision**, IET, v. 9, n. 6, p. 978–992, 2015.

OH, B.-S. et al. An analytic gabor feedforward network for single-sample and pose-invariant face recognition. **IEEE Transactions on Image Processing**, IEEE, v. 27, n. 6, p. 2791–2805, 2018.

OH, S.-K.; YOO, S.-H.; PEDRYCZ, W. Design of face recognition algorithm using PCA-LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks: Design and its application. **Expert Systems with Applications**, Elsevier, v. 40, n. 5, p. 1451–1466, 2013.

OH, S.-K.; YOO, S.-H.; PEDRYCZ, W. A comparative study of feature extraction methods and their application to P-RBF NNs in face recognition problem. **Fuzzy Sets and Systems**, Elsevier, v. 305, p. 131–148, 2016.

PAN, H.; ZHU, Y.; XIA, L. Fusing multi-feature representation and PSO-adaboost based feature selection for reliable frontal face detection. In: **Image Processing (ICIP), 2013 20th IEEE International Conference on**. Melbourne, VIC, Australia: IEEE, 2013. p. 2998–3002.

PARPINELLI, R. S.; LOPES, H. S. New inspirations in swarm intelligence: a survey. **International Journal of Bio-Inspired Computation**, Inderscience Publishers Ltd, v. 3, n. 1, p. 1–16, 2011.

PARPINELLI, R. S. et al. A review of techniques for on-line control of parameters in swarm intelligence and evolutionary computation algorithms. **International Journal of Bio-Inspired Computation (IJBIC)**, 2018. Forthcoming.

PASSINO, K. M. Biomimicry of bacterial foraging for distributed optimization and control. **IEEE control systems**, IEEE, v. 22, n. 3, p. 52–67, 2002.

PETERSEN, K. et al. Systematic mapping studies in software engineering. v. 8, p. 68–77, 2008.

PHAM, D. et al. The bees algorithm-a novel tool for complex optimisation. In: **Intelligent Production Machines and Systems-2nd I* PROMS Virtual International Conference (3-14 July 2006)**. Cardiff, UK: Science Direct, 2011.

PHILLIPS, P. J. et al. The FERET evaluation methodology for face-recognition algorithms. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 22, n. 10, p. 1090–1104, 2000.

PHILLIPS, P. J. et al. The feret database and evaluation procedure for face-recognition algorithms. **Image and vision computing**, Elsevier, v. 16, n. 5, p. 295–306, 1998.

PLICHOSKI, G. F.; CHIDAMBARAM, C.; PARPINELLI, R. S. A face recognition framework for illumination compensation based on bio-inspired algorithms. **7th Brazilian Conference on Intelligent Systems (BRACIS)**, In Press, 2018.

PLICHOSKI, G. F.; CHIDAMBARAM, C.; PARPINELLI, R. S. A self-adaptive face recognition framework based on differential evolution. **Swarm and Evolutionary Computation**, Elsevier, 2018. Submitted to.

PLICHOSKI, G. F.; PARPINELLI, R. S.; CHIDAMBARAM, C. Optimizing a homomorphic filter for illumination compensation in face recognition using population-based algorithms. **2017 Workshop of Computer Vision (WVC)**, IEEE, p. 78–83, 2017.

PLICHOSKI, G. F.; PARPINELLI, R. S.; CHIDAMBARAM, C. An adjustable face recognition system for illumination compensation based on differential evolution. **XLIV Latin American Computing Conference**, 2018. In Press.

PLICHOSKI, G. F.; PARPINELLI, R. S.; CHIDAMBARAM, C. A photometric-based illumination compensation approach for face recognition. **Anais do Computer on the Beach 2018**, 2018.

PLICHOSKI, G. F.; PARPINELLI, R. S.; CHIDAMBARAM, C. Swarm intelligence and evolutionary computation approaches for 2d face recognition: a systematic review. **Revista Brasileira de Computação Aplicada**, v. 10, n. 2, p. 2–17, 2018. ISSN 2176-6649. Disponível em: <http://seer.upf.br/index.php/rbca/article/view/8046>.

PRABHU, N. S. et al. Face recognition using eccentricity-range based background removal and multi-scaled fusion as pre-processing techniques. In: **Engineering (NUiCONE), 2013 Nirma University International Conference on**. Ahmedabad, India: IEEE, 2013. p. 1–6.

PRODOSSIMO, F. das C. et al. Proposta de uma metodologia para a construç ao de um banco de imagens faciais normalizadas. v. 5, n. 68, p. 41368, 2012.

RAO, N. M. et al. Face recognition using entropy based feature enhancement and parallel dual pose testing. In: **Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014 International Conference on**. Greater Noida, India: IEEE, 2014. p. 352–357.

RAO, R. Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. **International Journal of Industrial Engineering Computations**, v. 7, n. 1, p. 19–34, 2016.

RAO, S.; RAO, M. B. A novel triangular DCT feature extraction for enhanced face recognition. In: **Intelligent Systems and Control (ISCO), 2016 10th International Conference on**. Coimbatore, India: IEEE, 2016. p. 1–6.

RASHEDI, E.; NEZAMABADI-POUR, H.; SARYAZDI, S. GSA: A gravitational search algorithm. **Information Sciences**, v. 179, n. 13, p. 2232 – 2248, 2009. Special Section on High Order Fuzzy Sets.

RINKY, B. et al. DWT based feature extraction using edge tracked scale normalization for enhanced face recognition. **Procedia Technology**, Elsevier, v. 6, p. 344–353, 2012.

ROOHI, S. T. et al. Feature accentuation using uniform morphological correction as pre-processing technique for DWT based face recognition. **Procedia Technology**, Elsevier, v. 10, p. 690–698, 2013.

SAH, R. et al. Entropic-GWT based feature extraction and LBPSO based feature selection for enhanced face recognition. In: **Communications and Signal Processing (ICCSP), 2015 International Conference on**. Melmaruvathur, India: IEEE, 2015. p. 0180–0184.

SAMARIA, F. S.; HARTER, A. C. Parameterisation of a stochastic model for human face identification. In: **Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on**. Sarasota, FL, USA, USA: IEEE, 1994. p. 138–142.

SATTIRAJU, M. et al. Adaptive BPSO based feature selection and skin detection based background removal for enhanced face recognition. In: **Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on**. Jodhpur, India: IEEE, 2013. p. 1–4.

SCHEENSTRA, A.; RUIFROK, A.; VELTKAMP, R. A survey of 3D face recognition methods. In: **Audio-and Video-Based Biometric Person Authentication**. Berlin, Heidelberg: Springer, 2005. p. 325–345.

SHAN, S. et al. Illumination normalization for robust face recognition against varying lighting conditions. In: **Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on**. Nice, France, France: IEEE, 2003. p. 157–164.

SHANBHAG, S. S. et al. Face recognition using wavelet transforms-based feature extraction and spatial differentiation-based pre-processing. In: **Science Engineering and Management Research (ICSEMR), 2014 International Conference on**. Chennai, India: IEEE, 2014. p. 1–8.

SHANG, K. et al. A single gallery-based face recognition using extended joint sparse representation. **Applied Mathematics and Computation**, Elsevier, v. 320, p. 99–115, 2018.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality. **Biometrika**, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.

SHETTY, S. et al. Shift invariance based feature extraction and weighted BPSO based feature selection for enhanced face recognition. **Procedia Technology**, Elsevier, v. 10, p. 822–830, 2013.

SHIEH, M.-Y. et al. Applications of PCA and SVM-PSO based real-time face recognition system. **Mathematical Problems in Engineering**, Hindawi Publishing Corporation, v. 2014, 2014.

SIM, T.; BAKER, S.; BSAT, M. The CMU pose, illumination, and expression (PIE) database. In: **Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on**. Washington, DC, USA: IEEE, 2002. p. 53–58.

SOUMYA, S. et al. Face recognition using wavelet filtering and dual pose testing. In: **India Conference (INDICON), 2013 Annual IEEE**. Mumbai, India: IEEE, 2013. p. 1–6.

SPACEK, L. **University of Essex face database**. 2002.

STORM, R. Price, k.,". **Minimizing the Real Functions of the ICEC 96 contest by differential Evolution**, p. 842–844, 1997.

STORN, R.; PRICE, K. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. **Journal of global optimization**, Springer, v. 11, n. 4, p. 341–359, 1997.

SURABHI, A. et al. Background removal using k-means clustering as a preprocessing technique for DWT based face recognition. In: IEEE. **Communication, Information & Computing Technology (ICCICT), 2012 International Conference on**. Mumbai, India, 2012. p. 1–6.

THOMAZ, C. E.; GIRALDI, G. A. A new ranking method for principal components analysis and its application to face image analysis. **Image and Vision Computing**, Elsevier, v. 28, n. 6, p. 902–913, 2010.

TRINH, T. D.; KIM, J. Y.; NA, S. Y. Enhanced face recognition by fusion of global and local features under varying illumination. In: **IT Convergence and Security (ICITCS), 2014 International Conference on**. Beijing, China: IEEE, 2014. p. 1–4.

VALUVANATHORN, S.; NITSUWAT, S.; HUANG, M. L. Multi-feature face recognition based on PSO-SVM. In: **ICT and Knowledge Engineering (ICT & Knowledge Engineering), 2012 10th International Conference on**. Bangkok, Thailand: IEEE, 2012. p. 140–145.

VARADARAJAN, K. et al. Face recognition using block based feature extraction with CZT and goertzel-algorithm as a preprocessing technique. **Procedia Computer Science**, Elsevier, v. 46, p. 1458–1467, 2015.

VARUN, R. et al. Face recognition using hough transform based feature extraction. **Procedia Computer Science**, Elsevier, v. 46, p. 1491–1500, 2015.

VENKATESAN, S.; MADANE, S. S. R. Face recognition system with genetic algorithm and ant colony optimization. **International Journal of Innovation, Management and Technology**, IACSIT Press, v. 1, n. 5, p. 469, 2010.

VIDYA, V. et al. Face recognition using threshold based DWT feature extraction and selective illumination enhancement technique. **Procedia Technology**, Elsevier, v. 6, p. 334–343, 2012.

VISHWAKARMA, V. P. Illumination normalization using fuzzy filter in DCT domain for face recognition. **International Journal of Machine Learning and Cybernetics**, Springer, v. 6, n. 1, p. 17–34, 2015.

VORA, A. et al. Enhanced face recognition using 8-connectivity-of-skin-region and standard-deviation-based-pose-detection as preprocessing techniques. In: **Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014 International Conference on**. Greater Noida, India: IEEE, 2014. p. 364–369.

WECHSLER, H. et al. **Face recognition: From theory to applications**. Stirling, Scotland, UK: Springer Science & Business Media, 2012. v. 163.

WEN, G. et al. Improving face recognition with domain adaptation. **Neurocomputing**, Elsevier, v. 287, p. 45–51, 2018.

WEYRAUCH, B. et al. Component-based face recognition with 3D morphable models. In: **Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on**. Washington, DC, USA: IEEE, 2004. p. 85–85.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics bulletin**, JSTOR, v. 1, n. 6, p. 80–83, 1945.

XIAO, T. et al. Based on grid-search and PSO parameter optimization for support vector machine. In: **Intelligent Control and Automation (WCICA), 2014 11th World Congress on**. Shenyang, China: IEEE, 2014. p. 1529–1533.

YADAV, D. et al. Bacteria foraging fusion for face recognition across age progression. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops**. Portland, OR, USA: IEEE, 2013. p. 173–179.

YAJI, G. S. et al. DWT feature extraction based face recognition using intensity mapped unsharp masking and laplacian of gaussian filtering with scalar multiplier. **Procedia Technology**, Elsevier, v. 6, p. 475–484, 2012.

YANG, J. et al. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 39, n. 1, p. 156–171, 2017.

YANG, X.-S.; DEB, S. Cuckoo search via lévy flights. In: **Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on**. Coimbatore, India: IEEE, 2009. p. 210–214.

YIN, H.; FU, P.; SUN, Z. Face feature selection and recognition using separability criterion and binary particle swarm optimization algorithm. **Chinese Journal of Electronics**, v. 2, p. 025, 2014.

YIN, H. et al. Face feature selection with binary particle swarm optimization and support vector machine. **J. Inf. Hiding Multimed. Signal Process**, v. 5, n. 4, p. 731–739, 2014.

YOO, S.-H.; OH, S.-K.; PEDRYCZ, W. Design of face recognition algorithm realized with feature extraction from 2D-LDA and optimized polynomial-based RBF NNs. In: **IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint**. Edmonton, AB, Canada: IEEE, 2013. p. 655–660.

ZHANG, L.; YU, H.; HU, S. A new approach to improve particle swarm optimization. In: CANTÚ-PAZ, E. et al. (Ed.). **Genetic and Evolutionary Computation — GECCO 2003**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. p. 134–139.

ZHANG, W. et al. Improving shadow suppression for illumination robust face recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2018.

ZHANG, Y.; PENG, H. One sample per person face recognition based on particle swarm optimisation. **IET Signal Processing**, IET, v. 10, n. 2, p. 100–105, 2016.

ZHANG, Y.; WANG, S.; JI, G. A comprehensive survey on particle swarm optimization algorithm and its applications. **Mathematical Problems in Engineering**, Hindawi Publishing Corporation, v. 2015, 2015.

ZHAO, W. et al. Face recognition: A literature survey. **ACM computing surveys (CSUR)**, ACM, v. 35, n. 4, p. 399–458, 2003.

ZHAO, X. et al. Robust face recognition using the GAP feature. **Pattern Recognition**, Elsevier, v. 46, n. 10, p. 2647–2657, 2013.

ZHENG, W.-S.; LAI, J.-H.; YUEN, P. C. GA-fisher: a new LDA-based face recognition algorithm with selection of principal components. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 35, n. 5, p. 1065–1078, 2005.

ZOU, D.; ZHANG, H. Face recognition method based on 2DLDA and SVM optimated by PSO algorithm. **Advances in Intelligent Systems Research**, v. 130, 01 2016.

ANNEX A – WAVELET FAMILIES

Table 25 – Wavelet Families

bior1.1	bior1.3	bior1.5	bior2.2	bior2.4	bior2.6	bior2.8	bior3.1	bior3.3	bior3.5
bior3.7	bior3.9	bior4.4	bior5.5	bior6.8	coif1	coif2	coif3	coif4	coif5
coif6	coif7	coif8	coif9	coif10	coif11	coif12	coif13	coif14	coif15
coif16	coif17	db1	db2	db3	db4	db5	db6	db7	db8
db9	db10	db11	db12	db13	db14	db15	db16	db17	db18
db19	db20	db21	db22	db23	db24	db25	db26	db27	db28
db29	db30	db31	db32	db33	db34	db35	db36	db37	db38
haar	rbior1.1	rbior1.3	rbior1.5	rbior2.2	rbior2.4	rbior2.6	rbior2.8	rbior3.1	rbior3.3
rbior3.5	rbior3.7	rbior3.9	rbior4.4	rbior5.5	rbior6.8	sym2	sym3	sym4	sym5
sym6	sym7	sym8	sym9	sym10	sym11	sym12	sym13	sym14	sym15
sym16	sym17	sym18	sym19	sym20					