FEDERAL UNIVERSITY OF TECHNOLOGY - PARANÁ
GRADUATE SCHOOL OF ELETRICAL ENGINEERING AND APPLIED COMPUTER SCIENCE

CHIDAMBARAM CHIDAMBARAM

A CONTRIBUTION FOR SINGLE AND MULTIPLE FACES RECOGNITION USING FEATURE-BASED APPROACHES

DOCTORAL THESIS

CURITIBA

—June 2013

CHIDAMBARAM CHIDAMBARAM

A CONTRIBUTION FOR SINGLE AND MULTIPLE FACES RECOGNITION USING FEATURE-BASED APPROACHES

Thesis submitted to the Graduate School of Electrical Engineering and Applied Computer Science of the Federal University of Technology - Paraná in partial fulfillment of the requirements for the title of Doctor of Philosophy in Computer Engineering

Advisor: Dr. Heitor Silvério Lopes

Co-advisor: Dr. Hugo Vieira Neto

CURITIBA

—June 2013

Dados Internacionais de Catalogação na Publicação

C533 Chidambaram, Chidambaram

A contribution for single and multiple faces recognition using feature-based approaches / Chidambaram, Chidambaram. — 2013.

101f. : il. ; 30 cm

Orientador: Heitor Silvério Lopes.

Co-Orientador: Hugo Vieira Neto.

Tese (Doutorado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. Área de concentração: Engenharia de Computação. Curitiba, 2013.

Bibliografia: p. 91-96.

1. Biometria. 2. Percepção facial. 3. Colônia artificial de abelhas (ABC). 4. Reconhecimento de rosto humano (Computação). 5. Engenharia elétrica – Dissertações. I. Lopes, Heitor Silvério, orient. II. Vieira Neto, Hugo, co-orient. III. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. IV. Título.

CDD (22. ed.) 621.3

Biblioteca Central da UTFPR, Campus Curitiba

UTFPR
CAMPUS CURITIBA

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Câmpus Curitiba

CPGEI

Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial

Tese de Doutorado Nº. 89

"A Contribution for Single and Multiple Faces Recognition Using Feature-based approaches"

por

Chidambaram Chidambaram

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI, da Universidade Tecnológica Federal do Paraná – UTFPR, às 13h30min do dia 28 de junho de 2013, como requisito parcial à obtenção do título de Doutor em CIÊNCIAS - Área de Concentração: Engenharia de Computação. O trabalho foi aprovado pela Banca Examinadora composta pelos doutores:

Prof. Hugo Vieira Neto, Dr.
(Presidente - UTFPR)

Prof. Carlos Eduardo Thomaz, Dr.
(FEI)

Prof. Luiz Antônio Pereira Neves, Dr.
(UFPR)

Prof^a. Leyza Elmeri Baldo Dorini, Dr.
(UTFPR)

Prof. Gustavo Benvenutti Borba, Dr.
(UTFPR)

Visto da Coordenação:

Prof. Ricardo Lüders, Dr.
(Coordenador do CPGEI)

I dedicate this work to my family members who passed away during the period of my doctorate course:
My mother Muthulakshmi Chidambaram (2010)
My elder sister Rajalakshmi Chinniah (2012)
My brother Muthaiah Chidambaram (2012)
My cousin Chidambaram Ramasamy (2011)

ACKNOWLEDGEMENTS

I express my special thanks to:
- God and spirit guides who always provide support and guidance in achieving my goals in all moments of my life.
- My family, my wife and my son, for their inestimable support, motivation, comprehension, patience and encouragement for the past five years.
- Advisor Prof. Heitor Silvério Lopes, co-advisor Prof. Hugo Vieira Neto and Prof. Leyza Dorini for their support and guidance with patience in all moments of my research.
- My colleagues of the Bioinfo lab, UTFPR, César M. Vargas Benítez, Hugo Alberto Perlin, Flávio das Chagas Prodossimo, Marlon Subtil Marçal, Rafael Parpinelli, Ademir Gabardo, André Macário Barros, Wagner Rodrigo Weinert, Jonas Krause, Marlon Scalabrin, and Douglas Melchioretto mainly for their friendship and technical support on implementations and other related issues whenever I needed during my study.
- Universidade do Estado de Santa Catarina - UDESC for the financial support.
- My colleagues of Department of System Information for the general support.
- Coordinators of CPGEI, Prof. Humberto R. Gamba, Prof. Fábio Kurt Schneider and Prof. Ricardo Luders for the financial support whenever I needed to participate in conferences.
- Secretary of CPGEI, Terezinha Strapasson for her support regarding course matters.
- Prof. Luis Antônio Pereira Neves, Federal University of Paraná (UFPR), for his inestimable support and encouragement.

To my angels, my wife Niraimathy and my son Ramesh, no words could express my gratitude for all the sacrifices they have made during my study.

Compreender a natureza de cada ser humano, Conviver com todos praticando tolerância, paciência e compreensão, Viver sem egoísmo, ganância e orgulho, Ter fé e amor, e Solidarizar com o próximo são atitudes que levam ao caminho para a evolução do homem na terra.
.:.Chidambaram

RESUMO

CHIDAMBARAM, Chidambaram. A CONTRIBUTION FOR SINGLE AND MULTIPLE FACES RECOGNITION USING FEATURE-BASED APPROACHES. 101 f. Doctoral Thesis – Graduate School of Eletrical Engineering and Applied Computer Science, Federal University of Technology - Paraná. Curitiba, —June 2013.

Entre os sistemas de reconhecimento biométrico, a biometria da face exerce um papel importante nas atividades de pesquisa e nas aplicações de segurança, pois a face pode ser obtida sem conhecimento prévio de um indivíduo. Atualmente, uma grande quantidade de imagens digitais e sequências de vídeo têm sido adquiridas principalmente sob condições não-controladas, frequentemente com ruído, borramento, oclusão e variação de escala e iluminação. Por esses problemas, o reconhecimento facial (RF) é ainda considerado como uma área de pesquisa ativa e uma tarefa desafiadora. A motivação vem do fato que o reconhecimento de faces nas imagens com fundo complexo e em base de imagens faciais têm sido uma aplicação de sucesso. Portanto, o principal foco deste trabalho é reconhecer uma ou mais faces em imagens estáticas contendo diversos individuos e um individuo (face) em uma base de imagens com faces únicas obtidas sob condições diferentes. Para trabalhar com faces múltiplas, uma abordagem semi-supervisionada foi proposta baseada em características locais invariantes e discriminativas. A extração de características (EC) locais é feita utilizando-se do algoritmo *Speeded-Up Robust Features* (SURF). A busca por regiões nas quais as características ótimas podem ser extraídas é atendida através do algoritmo ABC. Os resultados obtidos mostram que esta abordagem é robusta e eficiente para aplicações de RF exceto para faces com iluminação não-uniforme. Muitos trabalhos de RF são baseados somente na extração de uma característica e nas abordagens de aprendizagem de máquina. Além disso, as abordagens existentes de EC usam características globais e/ou locais. Para obter características relevantes e complementares, a metodologia de RF deve considerar também as características de diferentes tipos e semi-globais. Portanto, a abordagem hierárquica de RF é proposta baseada na EC como globais, semi-globais e locais. As globais e semi-globais são extraídas utilizando-se de *Color Angles*(CA) e *Edge Histogram Descriptors*(EHD) enquanto somente características locais são extraídas utilizando-se do SURF. Uma ampla análise experimental foi feita utilizando os três métodos individualmente, seguido por um esquema hierárquico de três-estágios usando imagens faciais obtidas sob duas condições diferentes de iluminação com expressão facial e uma variação de escala leve. Além disso, para CA e EHD, o desempenho da abordagem foi também analisado combinando-se características globais, semi-globais e locais. A abordagem proposta alcança uma taxa de reconhecimento alta com as imagens de todas as condições testadas neste trabalho. Os resultados enfatizam a influência das características locais e semi-globais no desempenho do reconhecimento. Em ambas as abordagens, tanto nas faces únicas quanto nas faces múltiplas, a conquista principal é o alto desempenho obtido somente com a capacidade discriminativa de características sem nenhum esquema de treinamento.

Palavras-chave: Reconhecimento Semi-Supervisionado, Algoritmo ABC, Extração de características, Características Semi-globais, Reconhecimento Hierárquico, Variação de Iluminação

ABSTRACT

CHIDAMBARAM, Chidambaram. A CONTRIBUTION FOR SINGLE AND MULTIPLE FACES RECOGNITION USING FEATURE-BASED APPROACHES. 101 f. Doctoral Thesis – Graduate School of Eletrical Engineering and Applied Computer Science, Federal University of Technology - Paraná. Curitiba, —June 2013.

Among biometric recognition systems, face biometrics plays an important role in research activities and security applications since face images can be acquired without any knowledge of individuals. Nowadays a huge amount of digital images and video sequences have been acquired mainly from uncontrolled conditions, frequently including noise, blur, occlusion and variation on scale and illumination. Because of these issues, face recognition (FR) is still an active research area and becomes a complex problem and a challenging task. In this context, the motivation comes from the fact that recognition of faces in digital images with complex background and databases of face images have become one of the successful applications of Computer Vision. Hence, the main goal of this work is to recognize one or more faces from still images with multiple faces and from a database of single faces obtained under different conditions. To work with multiple face images under varying conditions, a semi-supervised approach proposed based on the invariant and discriminative power of local features. The extraction of local features is done using Speeded-Up Robust Features (SURF). The search for regions from which optimal features can be extracted is fulfilled by an improved ABC algorithm. To fully exploit the proposed approach, an extensive experimental analysis was performed. Results show that this approach is robust and efficient for face recognition applications except for faces with non-uniform illumination. In the literature, a significant number of single FR researches are based on extraction of only one feature and machine learning approaches. Besides, existing feature extraction approaches broadly use either global or local features. To obtain relevant and complementary features from face images, a face recognition methodology should consider heterogenous features and semi-global features. Therefore, a novel hierarchical semi-supervised FR approach is proposed based on extraction of global, semi-global and local features. Global and semi-global features are extracted using Color Angles (CA) and edge histogram descriptors (EHD) meanwhile only local features are extracted using SURF. An extensive experimental analysis using the three feature extraction methods was done first individually followed by a three-stage hierarchical scheme using the face images obtained under two different lighting conditions with facial expression and slight scale variation. Furthermore, the performance of the approach was also analyzed using global, semi-global and local features combinations for CA and EHD. The proposed approach achieves high recognition rates considering all image conditions tested in this work. In addition to this, the results emphasize the influence of local and semi-global features in the recognition performance. In both, single face and multiple faces approaches, the main achievement is the high performance obtained only from the discriminative capacity of extracted features without any training schemes.

Keywords: Semi-supervised Recognition, ABC algorithm, Feature Extraction, Semi-Global Features, Hierarchical Recognition, Illumination Variation

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

FR	Face Recognition
PCA	Principal Component Analysis
SIFT	Scale Invariant Feature Transform
SURF	Speeded-Up Robust Features
PSO	Particle Swarm Optimization
ABC	Artificial Bee Colony
MFR	Multiple Faces Recognition
SFR	Single Face Recognition
CA	Color Angles
EHD	Edge Histogram Descriptors
FT	Fourier Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
LBP	Local Binary Pattern
ACO	Ant Colony Optimization
PSO	Particle Swarm Optimization
ABC	Artificial Bee Colony
G	Global
SG	Semi-Global
L	Local

TABLE OF CONTENTS

1 INTRODUCTION

1.1 OVERVIEW

Many security systems are based on biometrics such as fingerprint analysis and iris or retinal scan. These types of methods depend on the collaboration of the individuals whose identity has to be verified. Iris recognition is one of the most accurate systems among biometrics, but its implementation cost is high. Fingerprints are reliable and simple to use, but they are not recommendable for non-collaborative individuals. On the other hand, among biometric recognition systems, face biometrics play an important role in research activities and security applications, since data can be acquired even without any knowledge of individuals. Among these methods, face recognition (FR) is non intrusive and relatively accepted by individuals (TOLBA; EL-BAZ; EL-HARBY, 2006; ABATE; NAPPI; RICCIO; SABATINO, 2007).

Although some works related to automatic machine recognition of faces date from the 1970s, it is still an active area that receives significant attention from both public and private research communities (TAN; CHEN; ZHOU; ZHANG, 2006). Human beings can recognize faces in cluttered scenes with relative ease, but machine recognition is much more difficult. During the 1980s, FR research remained stable and no prominent works were developed. Since the early 1990s, research interest in FR has grown very significantly as defined in the classical work of Chellappa and coworkers (1995). FR received attention from different areas, including image processing, pattern recognition, computer vision, artificial intelligence, computer graphics, neuroscience, robotics and evolutionary computing (YANG; KRIEGMAN; AHUJA, 2002; ZHOU; SCHAEFER, 2010). Hence, many researches have been focusing on FR through face detection from cluttered scenes and face retrieval from databases.

The main goal of the FR is to identify one or more faces from still images or video images of a scene or from a database of face images searching for an input face image. The main actions of FR involve detection, feature extraction from face regions, classification, retrieval and, finally, recognition (ZHOU; SCHAEFER, 2010). Consequently, FR through face detection from still images and face retrieval from databases have become common applications of image processing and analysis.

The success of FR methods depends on the extraction of stable features. The geometric feature-based and template matching methods are commonly used in many FR applications. The geometric feature-based method extracts information from local features such as eyes, nose and mouth. Template matching methods use the pixel information as features. However, features of these methods can be easily affected by image variations, for example, size, pose and lighting conditions (YANG; KRIEGMAN; AHUJA, 2002).

Variation of illumination is one of the prominent bottlenecks of FR. Inadequate lighting may produce images with variation in both brightness and contrast, affecting the recognition process. Many approaches have been proposed to handle these issues (NABATCHIAN; ABDEL-RAHEEM; AHMADI, 2011; ARANDJELOVIC; CIPOLLA, 2009; SOLAR; QUINTEROS, 2006; MATSUMOTO; SHIRAI; SHIMADA; SAKIYAMA; MIURA, 2006).

Images can be represented by one or more types of features extracted from color, edges and texture. Features can be extracted using a variety of methods which may be based on spatial or frequency filters, for example. In many cases, a large quantity of features are extracted, and the corresponding feature vector may have up to thousands of elements (PRIYA; R.S.RAJESH, 2012). To cope with this issue, it is common in the FR literature to reduce the dimension of the feature vector by using Principal Component Analysis (PCA). In this context, it is worth to obtain a feature set with few parameters so that it can be compact and representative of a face image. Moreover, the selection of appropriate types of features considering the different environmental conditions is a complex task in FR applications.

Color features can be globally represented by histograms which summarize the color distribution in an image. Hence, histograms serve as tools for color image analysis and for many spatial domain processing techniques (GONZALEZ; WOODS, 2009). Although histograms are invariant to rotation and translation under appropriate conditions, they cannot deal with illumination variations.

Edges represent an important feature of images. To handle the variations generated by different lighting conditions, most of the approaches often measure some image property that is at least insensitive to the variability in the imaging conditions. In this sense, edges can be one of the face image properties, although they do not contain all the useful information for recognition. Though there are many works dealing with illumination variation and compensation, early approaches concentrated their effort in handling the variability due to illumination by detecting edges, because edges tend to be less sensitive to different lighting conditions (BELHUMEUR; KRIEGMAN, 1998). In addition to the definition of types of features, the extraction of relevant features and their representation are some other issues in FR.

In the FR context, it is essential to determine relevant features to be considered in the comparison between the face object image or query (input) image and faces in still images with multiple faces or single face images. At the same time, the extraction of invariant features that represent faces are important to overcome illumination variations and other related issues. Some works, for example, Lowe (2004), Mikolajczyk e Schmid (2004), and Trujillo e Olague (2006), have focused on extracting local image features called interest points. Interest points have a compact form of representing distinctive and invariant regions through their own descriptors and provide efficient indexing and retrieving of images (MIKOLAJCZYK; SCHMID, 2004). Interest point detectors have been mainly applied for object recognition and other related tasks (PEREZ; OLAGUE, 2009; LOWE, 2004; AZAD; ASFOUR; DILLMANN, 2009).

Among many image interest point descriptors, two can be mentioned as most known recently: Scale Invariant Feature Transform (SIFT) (LOWE, 2004) and Speeded-Up Robust Features (SURF) (BAY; ESS; TUYTELAARS; GOOL, 2008). SURF is considered as a scale and rotation-invariant detector and descriptor. SURF is considered less sensitive to the presence of noise in images and outperforms SIFT. Another major advantage of SURF is that it requires low computation time to detect and describe the interest points in images in comparison to SIFT (BAY; ESS; TUYTELAARS; GOOL, 2008). Though the different types of features that are invariant to image variations can contribute to the performance of FR, the extraction of them just as global features may not be sufficient to find the discriminative and representative features.

From the early to most recent studies on FR, many attempts have been made using either global or local features of images (SINGH; WALIA; MITTAL, 2012). Global features describe the image as a whole while local features represent small parts of the image. Likewise, global features have the potential of generalizing an entire image which may provide some guidelines for class discrimination purposes. On the contrary, local features obtain local information at multiple points or some interior parts of face images such as eyes, nose and mouth (AMARAL; THOMAZ, 2012). Since both global and local features are generally obtained in a distinct way, they can also provide different types of information (LISIN; MATTAR; BLASCHKO; LEARNED-MILLER; BENFIELD, 2005). Likewise, similar to global and local features, semi-global features (features extracted from regions that have a size between local and global regions as defined in EHD) can also provide some different information about face images. However combining local information with global information may aid in object recognition process (TAN; CHEN; ZHOU; ZHANG, 2006). According to Tan and colleagues, works based on hybrid methods are relatively few and still more research effort could be taken in this direction. Although, the features and its extraction methods play an important role in FR applications, in still images with multiple faces obtained under different conditions, the detection and extraction of invariant features may need some additional effort from search algorithms.

Traditional search algorithms are computationally expensive. However, many real-life FR applications require fast and efficient search and matching algorithms (PAWAR; TALBAR, 2009). Metaheuristic optimization algorithms, such as those from the Swarm Intelligence area, were successfully applied to the FR problems. Several successful optimization algorithms have been applied to FR purposes such as Particle Swarm Optimization (PSO) (PERLIN; LOPES; CENTENO, 2008; SUGISAKA; FAN, 2005) and Artificial Bee Colony (ABC) (CHIDAMBARAM; LOPES, 2010) algorithms.

Based on the context exposed in this section, the focus of this work is directed to propose novel FR approaches for still images obtained under varying conditions, overcoming the issues associated with images. Feature extraction methods and types of features play an important role in any FR process. Therefore, this proposal is based on the fact that some alternative approaches with different types of features and complementary information obtained from these features can somehow contribute to improve the performance of FR methods.

1.2 MOTIVATION

During the recent years, motivated by the presence of crimes, violence and other illegal activities, both public and private organizations are investing in surveillance systems. Due to the development of new technologies, nowadays, a huge amount of still images are available and were typically acquired under different imaging conditions. These conditions usually include noise, blur, pose changes, occlusion, scale and illumination variation. Consequently, recognizing faces from these images under those uncontrolled conditions becomes a challenging task. Additionally, other related issues such as face expression, hair style, cosmetics and aging still makes the face recognition problem more complex. Hence, many researches have been focusing on how to make full automatic recognition of human faces on still images and video sequences. According to the document released by the National Technology and Science Council (NSTC, 2006), even though many advancements have been made in FR, the need for systems with higher accuracy still remains.

According to the Face Recognition Vendor Test (FRVT 2006) and the study conducted by Beveridge and coworkers (BEVERIDGE; S.BOLME; A.DRAPER; GIVENS; LUI, 2010), among the factors that affect recognition performance, changes in illumination appears on the top of the list. Therefore attempts to construct a FR methodology that can minimize the effect illumination variations becomes a very active and promising research area to improve the performance rate.

In still images with multiple faces, face images can be found under different conditions with complex background and variations in the conditions of acquisition. Therefore, FR in such images becomes a complex problem and it still remains as an open research area.

FR approaches are mostly based on supervised training schemes. Therefore, in this work, a semi-supervised learning approach for images with multiple faces is proposed. This is one of our main contributions which developed in the present research. Furthermore, object recognition problems and some FR tasks are treated as an optimization problem and can be solved using swarm intelligence approaches. In this way, one of the motivations is to propose the use of the ABC algorithm in the searching process in still images with multiple faces.

In the literature, though many global and local methods for face recognition were discussed in recent years, there is still space for development of efficient approaches suitable for most of the image conditions. In this sense, one of the motivations comes from the study of how the different features are extracted from the images and how they can influence on the recognition performance. Additional motivation comes from the possibility of solving the single FR using only face image features without any application of learning approaches.

1.3 OBJECTIVES

1.3.1 GENERAL OBJECTIVE

The general objective of this work is to propose novel approaches to recognize faces in still images with multiple faces and in single face images under varying conditions using different types of features.

For multiple faces recognition (MFR), the general objective is to propose semi-supervised approach for FR, treating it as an optimization problem.

For single faces recognition (SFR), the general objective is to propose a single-stage and a three-stage hierarchical FR approach.

1.3.2 SPECIFIC OBJECTIVES

The MFR approach includes the following specific objectives:

- Construction of an image database with multiple faces under different illumination conditions and with image variations such as noise, blur, scale, occlusion and inclination (tilted head).
- Develop an semi-supervised approach for MFR.
- Treat the MFR as an optimization problem under different images conditions and solve it using the ABC algorithm.
- Investigate the influence of illumination variation in still images with multiple faces and different feature extraction methods.

The SFR approach includes the following specific objectives:

- Construction of an image database with single faces under different image conditions such as face expression (smiling), different illumination conditions, lateral face illumination and scale.
- Construction of image base features using the feature extraction methods such as color angles (CA), edge histogram descriptors (EHD) and SURF, and by extracting from global image and uniform subimages combination.
- Evaluate the recognition performance of FR methods independently (single-stage) using the image database.
- Study the recognition performance of FR methods under different features combinations using global, semi-global and local features.
- Construct a three-stage hierarchical SFR approach using the feature-extraction methods (color, edges and interest points).
- Compare the performance of SFR methods individually against hierarchical SFR approach based on recognition performance and computational effort.

1.4 STRUCTURE OF THE THESIS

The present work is divided into five chapters. Besides the general view of FR related to the problem, the first chapter presents the general objective, specific objectives, and finally, the motivations of this work. The second chapter covers the theoretical foundations focusing mainly on methods for face recognition and feature extraction such as EHD, CA and SURF, and also the ABC algorithm. In addition to these topics, it also discusses the other related works that are applied to solve the FR problem such as swarm intelligence, interest point detection and edge histogram descriptors. The methodology of this work is detailed in the third chapter. The topics regarding multiple faces recognition using ABC algorithm, heterogenous features extraction and their combination and hierarchical approach for single faces recognition are discussed. The fourth chapter describes the experiments and their results. A discussion on results is given in the same chapter. Finally, the last chapter summarizes the conclusions of the work. It also points out the future works and directions of research.

1.5 MAIN CONTRIBUTIONS

The main contributions of this work are defined as follows:

- Construction of novel approaches to recognize faces in still images with multiple faces and single faces under varying conditions;
- Treat the FR in still images as an optimization process using a swarm intelligence algorithm;
- Propose a general method for extracting features as global, semi-global and local and their combinations in SFR;
- Propose a hierarchical three-stage SFR with the proposed method of features extraction;
- Develop a hierarchical three-stage SFR and MFR as a semi-supervised approach with the aid of discriminative power of extracted features;

.

2 THEORETICAL FOUNDATION

2.1 FACE RECOGNITION

FR systems are generally classified into two categories: face verification (1:1) and face identification (1:n). While face verification performs the matching between a face image and a template face image, face identification aims to match a query face image against all template images in a database (ABATE; NAPPI; RICCIO; SABATINO, 2007). In other words, the latter consists in a one to n image matching process and the former in a one to one image match.

Based on the classification of Brunelli and Poggio (BRUNELLI; POGGIO, 1993) on FR methods and Yang and colleagues (YANG; KRIEGMAN; AHUJA, 2002) on face detection methods, FR methods can be widely classified into three main categories: Feature-based methods, template matching and appearance-based methods. Among the three main categories of FR methods, the first two methods have been widely used since the beginning, more specifically, using geometric-based features (BRUNELLI; POGGIO, 1993).

The geometric feature-based methods extract information from local features such as eyes, nose and mouth. Template matching methods are based on the matching between a predefined parameterized face template and an image containing a face. One of the classical works based on template matching and geometric feature-based matching was done by Brunelli and Poggio (BRUNELLI; POGGIO, 1993) in 1993. They performed experimental tests using the eyes, nose, mouth and a whole face templates in the same sequence and concluded that the single feature is powerful. Specific features of face images can also be used as a template to study the recognition performance (instead of whole face image). In this way, they extracted 35 geometrical facial features. Besides the location of the basic features (eyes, mouth, nose and eyebrows) and face outline. In addition to this, they also considered additional measures, such as eyebrow thickness and vertical position at the eye center position, eleven radii describing the chin shape and nose vertical position and width.

Appearance based methods can be defined as the projection of images into a lower dimensional sub-space before classifying them using a distance or similarity measure (STRUC;

PAVESIC, 2009). These methods can be applied either to the whole face or to a specific region in which recognition is performed using low-dimensional representation. Appearance based methods can be divided into two categories (SHAN; CAO; GAO; ZHAO, 2002): holistic appearance-based methods and analytic local-feature based methods.

For holistic appearance methods, FR is done by using the global features of the face image as input. It includes commonly used approaches such as Eigenfaces (TURK; PENTLAND, 1991) (based on Principal Component Analysis (PCA) (KRIBY; SIROVICH, 1990)), Fisherfaces (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) (based on Linear Discriminant Analysis (LDA)) , Singular Value Decomposition (SVD) (ZHANG; CHEN; ZHOU, 2005) and most of the neural networks (NN) based methods (ER; CHEN; WU, 2005).

PCA was first applied to represent pictures of human faces by Kirby and Sirovich (KRIBY; SIROVICH, 1990). Also known as Karhunen-Loeve expansion, It is a classical feature extraction and data representation technique applied in the pattern recognition and computer vision areas (YANG; ZHANG; FRANGI; YANG, 2004). PCA is the basis for the eigenfaces approach (TURK; PENTLAND, 1991). In summary, the PCA-Eigenfaces approaches uses the available images as a training set and tries to explore the fact that each face is a variation of a mean face.

Analytic methods, on the other side , take the advantage over the fact that all human faces share some similar topological structure. The most known methods are Local Feature Analysis (LFA) (FAZL-ERSI; TSOTSOS, 2009) and Elastic Bunch Graph Matching (EBGM) (WISKOTT; FELLOUS; N.KUIGER; MALSBURG, 1997). LFA represents the faces through a combination of local features which are described by a set of Gabor wavelet coefficients. EBGM uses a graph representation, where the nodes located at fiducial points are connected by labeled edges (FAZL-ERSI; TSOTSOS, 2009).

FR has to deal with large within-class variations caused due to the lighting conditions and different poses. To compensate such variations, it is desirable to incorporate them into the process of feature extraction. This was done by Etemad and Chellappa (ETEMAD; CHELLAPPA, 1997) through the LDA approach for human faces. Using this approach, they evaluated the discriminative potential of different facial features. In other words, human faces were analyzed in the spatial and the frequency domains. LDA is a statistical approach based upon a discriminative criterion that tries to maximize the between-class variance and minimize within-class variance of the scatter (BELHUMEUR; HESPANHA; KRIEGMAN, 1997).

Experimental studies have shown that image variations due to illumination and viewing direction influence the recognition performance much more than the variations due to the change in face identity (BELHUMEUR; KRIEGMAN, 1998). It is important to mention that all faces have the same common features with slight differences in the position, size, shape and color (KARUNGARU; FUKUMI; AKAMATSU, 2004), which can be disturbed by the mentioned variations. Over the last two decades, many works have been proposed to make the face recognition systems invariant to some of these problems.

In recent years, more generally, FR works are based on features-based method and appearance-based methods. However, the image features such as color, texture and edges are extracted using features-based methods rather geometrical-features. These features are sometimes extracted through image transformation applying spatial or frequency filters. From the extracted features space, invariant and discriminative features are obtained for FR tasks.

Color features are generally considered as a powerful features for image matching and other related purposes. For example, skin color information is used to locate face regions by separating them from complex backgrounds. This procedure can probably reduce the computational effort spent to detect faces in complex background regions (LIN, 2007). Using skin-color information, facial regions and specific facial features can be obtained. However, under different lighting conditions, skin color models may not work effectively. This problem can be overcome to a certain level by using color descriptors (MANJUNATH; OHM; VASUDEVAN; YAMADA, 2001) and illumination compensation techniques (ARANDJELOVIC; CIPOLLA, 2009). The color descriptors can capture the spatial distribution of the colors in a compact form which can be effective in search and retrieval applications. In this context, the application of complementary features of color spaces may improve the face recognition performance and fusing features across color spaces can also enhance the discrimination power (LIU; LIU, 2008).

Similar to features extracted from colors, edges also serve as an important feature to construct descriptors, for example, using histograms (MANJUNATH; OHM; VASUDEVAN; YAMADA, 2001) as in the case of edge histogram descriptors. In FR applications Hsu, Wang, Wang, Tseng e Tang (2010) applied Canny edge detectors to find the contours of faces. Feature extraction methods based on edge detection can explore some specific facial characteristics, such as eyes, mouth, nose, eyebrows, lips, chin, ears, cheek and face outline. However, among these facial features, the eyes, edges and lips are considered as primary features (KARUNGARU; FUKUMI; AKAMATSU, 2004). The edges, as primary features, certainly can be useful to construct descriptors of any image, for example, face.

Extracting features by preprocessing and segmentation was a standard approach. However, after the year 2000, instead of preprocessing and segmentation, it has become common that features are extracted from small patches around interest point and represented by descriptors. The interest points are generally regions in which some distinct features can be found. Since their development, interest points have been mainly applied for object recognition and other related tasks (TRUJILLO; OLAGUE, 2008; AZAD; ASFOUR; DILLMANN, 2009; BAY; ESS; TUYTELAARS; GOOL, 2008). In this way, face processing tasks using interest point detectors seems to be a promising area (FERNANDEZ; VICENTE, 2008; BAY; ESS; TUYTELAARS; GOOL, 2008). Several interest point detectors and descriptors have been proposed for features extraction such as SURF and SIFT.

Image transformation is a process of transforming images from the space domain to a transform domain. Basic image processing approaches or spatial domain techniques, for example, affine transformations, operate directly manipulating the pixels of an image. But, in some cases, operations are realized on frequency domain as in Fourier transform (FT) rather on image pixels. By image transformation, image feature can be analyzed from different angles and this can make the image processing and recognizing tasks may become simple and effective (LIHONG; YING; YUSHI; CHENG; XILI, 2009). In fact, spatial domain techniques are computationally more efficient and require less resources to implement (GONZALEZ; WOODS, 2009).

Based on the fact that the image transformation techniques including image-filtering techniques may improve the face recognition performance, many works have been presented in recent years (ARANDJELOVIC; CIPOLLA, 2009). These techniques are high-pass filters, directional derivatives, Laplacian-of-Gaussian filters, edge-maps and wavelet-based filters (GONZALEZ; WOODS, 2009). Besides these techniques, many image transformation algorithms can be found in the literature which commonly include DFT (Discrete Fourier Transform), DWT(Discrete Wavelet Transform), Walsh Transform, HT (Hadamard Transform), and KL transform (Karhunen-Loeve Transform) (LIHONG; YING; YUSHI; CHENG; XILI, 2009).

As discussed in this section, three feature extraction methods are explained in the following sections. Besides, swarm intelligence algorithm ABC is also presented in the subsequent sections.

2.2 FEATURE EXTRACTION METHODS

2.2.1 INTEREST POINT DETECTOR AND DESCRIPTOR - SURF

Object recognition using correlation-based methods can be unreliable where object pose and environment illumination are not tightly controlled. Instead of matching simply all the pixels, the alternative could be to match some features identified from invariant image locations. One important step is the identification of local image features which are invariant to image scaling, translation, rotation and illumination. In this context, interest points can be an alternative way of describing image features like color, texture and shape. In many recent computer vision applications, distinctive and representative regions of images are identified using interest points. In order to locate and select such regions of interest, different types of detectors are applied on images to select the pixels of distinctive values. Since its development, interest points have been mainly applied for object recognition and other related tasks (PEREZ; OLAGUE, 2009; LOWE, 2004; AZAD; ASFOUR; DILLMANN, 2009).

A interest point detector is an algorithm that uses an image as an input and outputs the image with a set of interest points that can be identified with high repeatability in location (ZULIANI; M.KENEDY; MANJUNATH, 2004). Object recognition may be a successful process only if it is possible to find some distinctive image features among many alternative objects in cluttered real scenes (LOWE, 1999). Hence, this is the one of the major objective of interest point detectors. Nowadays, interest points detection can be considered as an appropriate method for solving recognition problems in the computer vision and related areas.

Pixels are correlated spatially in the 2-D intensity array of images and the information carried out by individual pixels alone is not representative. Therefore more information can be obtained from neighborhood pixels. Irrelevant information and spatial redundancy are present in the most of the pixels of the 2-D intensity arrays. Both of them can be reduced by transforming the 2-D intensity array in to a more efficient ”non-visual”representations (GONZALEZ; WOODS, 2009). Interest points can be associated to these representations. Interest points can be generally characterized in several ways. They can be defined as a set of image pixels that have high level of variation in reference to a predetermined local measure (TRUJILLO; OLAGUE, 2006). They are salient regions that are highly distinctive with local minimum or maximum intensity in an image (LOWE, 2004). Comparing corners with interest points, it can be said that the corners are mostly an intersection of two edges meanwhile interesting points are points in an image which have a well-defined position. Compared to low-level features like color, interest points are considered as more stable and reliable (PIMENOV, 2009). Object recognition can be considered as an main application of interest point (LOWE, 1999; AZAD; ASFOUR; DILLMANN, 2009) detectors. Likewise, since the development of interest point detectors and invariant descriptors, FR and related tasks have also become one of its application (ASBACH; P.HOSTEN; UNGER, 2008; FERNANDEZ; VICENTE, 2008).

Most of the current interest point detector algorithms are manually designed using different image processing techniques and on the current understanding of how humans recognize objects. Trujillo and Olague (TRUJILLO; OLAGUE, 2008) proposed a novel evolutionary approach using Genetic Programming as an optimization problem which results in an automatic generation of interest point operators. GP was used to combine low level image operators in order to segment an image. At the end of the optimization process, the final image that is obtained as an output is used to detect interesting points (TRUJILLO; OLAGUE, 2008). According to the proposed approach, the evolved interest point operators should fulfill three properties: Global separability, High information content and stability.

2.2.1.1 INTEREST POINT DETECTORS

To solve the object recognition problem, one way of detecting feature location is the use of a corner detector, which is a detector of local image regions which have a high degree of variation in all directions. Corners can be L-corners, T-junctions and Y-junctions and other locations with any significant 2D texture (SCHMID; MOHR; BAUCKHAGE, 2000). Corner detectors when operating directly on intensity images are also referred to as interest point detectors (TRUJILLO; OLAGUE, 2006). Although corner detectors are widely used, they are not usually as robust as like interest point detectors. For example, the Harris corner detector can examine an image at only a single scale, i.e, it is very sensitive to size change (LOWE, 1999).

The main interest point detectors include: Harris (HARRIS; STEPHENS, 1988) (invariant to rotation), Harris-Laplace (invariant to rotation and scale changes) and Harris-Affine (MIKOLAJCZYK; SCHMID, 2005) (invariant to affine image transformations), and DoG (LOWE, 2004) (invariant to rotation and scale changes). There are also many interest point descriptors, but, two of them can be mentioned as most known recently: Scale Invariant Features Transform (SIFT) (LOWE, 2004) and Speeded-Up Robust Features (SURF) (BAY; ESS; TUYTELAARS; GOOL, 2008). Most of the detectors finally generate descriptor vectors which contain the information regarding the neighborhood of every interesting point in an image. Both SIFT and SURF-based methods are used to detect interest points but the construction and implementation of these detectors follow different methods (BAY; ESS; TUYTELAARS; GOOL, 2008; LOWE, 2004).

In 2008, Bay and colleagues (BAY; ESS; TUYTELAARS; GOOL, 2008) developed a novel scale and rotation-invariant detector and descriptor, SURF . This approach uses a very basic Hessian matrix approximation, which is also used in SIFT and other approaches, to detect interest points by selecting distinctive locations like corners, blobs and T-junctions. Another important aspect of this work is the detection of interest points at different scales. Instead reducing the image size as a pyramid (LOWE, 1999), scale space is analyzed by up-scaling the filter size as shown in Figure 1 (a). To implement this scheme, scale space is divided into octaves which are a series of filter response maps. Octaves are obtained by convolving the same input image with a filter of increasing size. In this scheme, due to the use of integral images (VIOLA; JONES, 2001), the computational cost is constant.

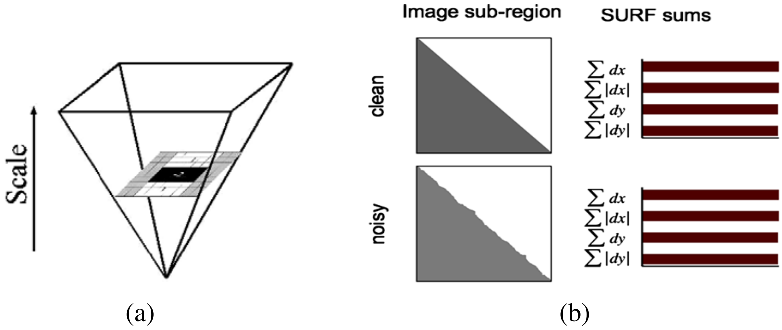


Figure 1: (a) Analyzing the scale space by up-scaling the filter size using a pyramid scheme (b) Image sub-region and its sum of intensity patterns

According to Bay and colleagues, although SURF can be similar in concept as SIFT, SURF is less sensitive to noise and outperforms SIFT. It happens because of the global integration of gradient information obtained from sub-region instead of individual gradients as in the case of SIFT. This is illustrated in Figure 1(b). Another major advantage of SURF is that it requires low computation time to detect and describe the interest points in comparison with other two schemes SIFT (LOWE, 2004) and GLOH (MIKOLAJCZYK; SCHMID, 2005). Experiments conducted using this approach indicate that SURF is suitable for object recognition and image retrieval tasks.

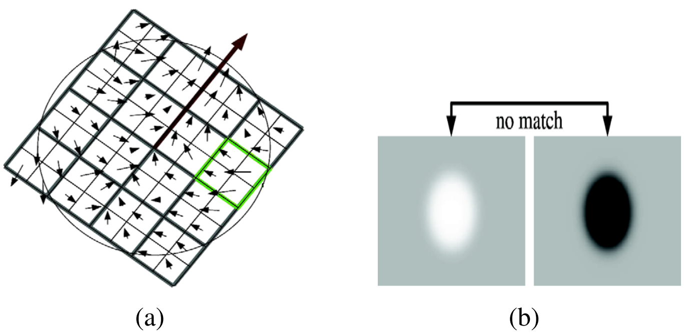


Figure 2: (a) Descriptor construction using oriented quadratic grid with 4 x 4 square sub-region and Haar wavelet responses ($\sum d_x, \sum d_y, \sum \mid d_x \mid, \sum \mid d_y \mid$) around an interest point (b) Interest point matching using contrast type

2.2.1.2 DESCRIPTORS

Image descriptors are descriptions of the connection between the pixels that are identified as distinctive and stable by the algorithm of a detector. They generally describe the features of an image like shape, color, texture and also motion. But, the descriptors of interest points are represented by vectors which contain information of the interest point neighborhood. SURF builds a descriptor vector of 64 dimension which is obtained concatenating all 4 x 4 sub-regions of 4 dimensional vectors of underlying intensity structure as shown in Figure 2(a). To build its descriptor, the SURF extracts from the distribution of first order Haar wavelet responses in x and y direction (BAY; ESS; TUYTELAARS; GOOL, 2008). For each sub-region the wavelet responses in the horizontal direction (dx) and in the vertical direction (dy) are accumulated and it forms the first set of entries in the feature vector. Furthermore, indexing during the matching stage is based on the sign of the Laplacian . This avoids matching of features with different type of contrast which is demonstrated in Figure2(b). Hence, the way the distribution of the intensity content within interest point neighborhood is obtained and described by the SURF descriptor, reduces the time for feature computation and matching, and also increases robustness. Although SIFT is more popular, the performance of SURF is equal or better than SIFT and its computational efficiency is significantly better than SIFT (PIMENOV, 2009).

2.2.1.3 INTEREST POINT EVALUATION CRITERIA

The main task in FR is to find out the similar features between two face images. Generally features should have some specific properties that can be used in matching images, for example, robustness and distinctiveness. Robustness refers to the invariant features to illumination, scale and pose variations and distinctiveness indicates the uniqueness of features. Large number of features can be extracted from face images using different algorithms. The main fact is that such features should be highly distinctive and provide a basis for the recognition task. Interest points can also be treated in the same way. They should satisfy basically three important properties: global separability between extracted points, high information content when compared to other pixels, and stability under certain types of image transformations (TRUJILLO; OLAGUE, 2008).

The global separability defines that the interest points should be extracted from different parts of an image. In other words, they should be found spread out in the entire space of an image. Information content measures the distinctiveness of these interest points and thus, the distribution of the descriptors indirectly. If the interest points are cluttered together, then it implies that the descriptors may be similar and thus, it will result in low information content(SCHMID; MOHR; BAUCKHAGE, 2000). From this brief discussion, it can be said that the matching will be successful only if the descriptors have high information content and if they are distinctive.

Besides the properties like high information content and global separability, the stability of interest points is the most important property to achieve high rate of correct matching. Repeatability rate is the only measure of stability which is strongly accepted as a standard computer vision performance metric for interest points (TRUJILLO; OLAGUE, 2006).

Repeatability is defined by the image geometry. Measurements of repeatability will quantify the number of repeated points detected under varying conditions such as image rotation, scale change, variation of illumination, presence of noise and view point change. The percentage of detected points that are repeated in both images is defined as the repeatability rate. Repeatability criteria is valid only for planar scenes in which the geometric relation between two images is completely defined (SCHMID; MOHR; BAUCKHAGE, 2000).

In summary, the percentage of points repeated in the two images being compared is defined as the repeatability rate. A point is considered repeated if it lies in the same coordinates on both images. Due to the several variations or transformations present in real-world conditions, a point is in general not detected exactly at the same position, but in some neighborhood. Thus, an acceptable error needs to be established when measuring the distance between the coordinates of two images.

Hence, the set of repeated interest points on images I_j and I_k, denoted by $Rip_{j,k}$, is defined as:

$$Rip_{j,k} = \{\mathbf{x}_i \mid \sqrt{(\mathbf{x}_i^j - \mathbf{x}_i^k)^2} < T_{CDE}\} \tag{1}$$

where $\mathbf{x}_i^n = (x_i^n, y_i^n)$ denotes the i-th coordinate (x, y) in the image n and T_{CDE} represents the acceptable distance error between the coordinates of interest points on different images (Coordinate Distance Error - CDE).

If the point is classified as repeated, then an acceptable distance error for the associated descriptors also needs to be defined:

$$RIP_{j,k} = \{\mathbf{x}_i \in Rip_{j,k} \mid \sqrt{\sum_{i=1}^{n} (\mathbf{d}_i^j - \mathbf{d}_i^k)^2} < T_{DDE}\} \tag{2}$$

where d_i^n denotes the i-th position of the descriptor vector related to the interest point $\mathbf{x}_i$ of image n, and T_{DDE} represents the admissible distance error between two descriptor vectors (Descriptor Distance Error - DDE).

The repeatability rate, R, of interest points extracted from two images, Im_j and Im_k, is defined by the following equation:

$$R = \frac{RIP_{j,k}}{\min(IP_j, IP_k)} \tag{3}$$

where $RIP_{j,k}$ denotes the repeated interest points obtained by Equation 1, and NI_j and NI_k represent the total of number of interest points detected on images Im_j and Im_k, respectively. The image with minimum number of interest points is considered since the number of detected points may be different for the two images.

2.2.2 COLOR ANGLES

The color angles approach is initially proposed by Finalyson and colleagues (FINLAYSON; CHATTERJEE; FUNT, 1996). Their main goal was to develop a color based image descriptor that is concise, expressive and illumination variant. In addition to this, they aimed to construct a descriptor with very few parameters so that it can be useful for computationally intensive algorithms.

In images, illumination variation can be treated as a linear transform of image colors (FINLAYSON; CHATTERJEE; FUNT, 1996). With the presence of noise, the representation of this transform may not be stable and expressive. However, a scene viewed under two different illuminants can be connected by 3 simple scale factors. In this case, each pixel in the first image (R_i, G_i, B_i) becomes (s_1R_i, s_2G_i, s_3B_i) where s_1, s_2 and s_3 are scalars. These scalars represents the illuminant-variant information between two images seen under same view. Hence, the images that differs only in terms of the scene illuminant can be related by a simple 3-parameter diagonal matrix (FINLAYSON; CHATTERJEE; FUNT, 1996). If an image band is represented as a vector in a high-dimensional space, when the illumination changes, then the vector becomes longer or shorter but its orientation remains unchanged. Hence, under a diagonal model of illuminant change, the 3 angles between the different bands of an image can define the illuminant invariant relation.

The investigation of colors under a changing illuminant indicates that likely illuminant changes can be defined as a restricted subset of linear transforms. From this observation, we can extract useful illuminant-invariant statistics from color distributions. Hence the color distributions represented by color angles encode important low-order statistical information (FINLAYSON; CHATTERJEE; FUNT, 1996). One way of describing the color distributions using a statistical measure is by a mean image color which is defined as follows:

$$\overline{r} = \frac{1}{M}\sum_{i=0}^{M-1}(r_i) \tag{4}$$

$$\overline{g} = \frac{1}{M}\sum_{i=0}^{M-1}(g_i) \tag{5}$$

$$\overline{b} = \frac{1}{M}\sum_{i=0}^{M-1}(b_i) \tag{6}$$

where M represents the number of pixels and r_i, g_i and b_i are red, green, and blue pixels values of an image respectively.

In this work, the three color angles are calculated based on the approach proposed in (NETO, 2011), which uses a simplified form of (FINLAYSON; CHATTERJEE; FUNT, 1996). The relation between three channels R, G and B is represented by three color angles ϕ_{rg}, ϕ_{gb} and ϕ_{rb}. The first step to compute the color angles consists on the calculation of zero-mean color vectors r_0, g_0 and b_0 as defined by the following equations:

$$\mathbf{r}_0 = \mathbf{r} - \overline{r} \tag{8}$$

$$\mathbf{g}_0 = \mathbf{g} - \overline{g} \tag{9}$$

$$\mathbf{b}_0 = \mathbf{b} - \overline{b} \tag{10}$$

where $\overline{r}$, $\overline{g}$ and $\overline{b}$ are the average pixel values of the color channels R, G and B, respectively (Equation 4). In the following step, the zero-mean color vectors are normalized and they are represented by $\mathbf{r}_N$, $\mathbf{g}_N$ and $\mathbf{b}_N$:

$$\mathbf{r}_N = \frac{\mathbf{r}_0}{||\mathbf{r}_0||} \tag{12}$$

$$\mathbf{g}_N = \frac{\mathbf{g}_0}{||\mathbf{g}_0||} \tag{13}$$

$$\mathbf{b}_N = \frac{\mathbf{b}_0}{||\mathbf{b}_0||} \tag{14}$$

To calculate the color angle between two color channels, for example, R and G, based on the a geometrical assumption, we calculate the inverse cosine of dot products of the normalized vectors of two color channels:

$$\phi_{rg} = \arccos(\langle \mathbf{r}_N, \mathbf{g}_N \rangle) \tag{16}$$

$$\phi_{br} = \arccos(\langle \mathbf{b}_N, \mathbf{r}_N \rangle) \tag{17}$$

$$\phi_{gb} = \arccos(\langle \mathbf{g}_N, \mathbf{b}_N \rangle) \tag{18}$$

2.2.3 EDGE HISTOGRAM DESCRIPTOR (EHD)

Edge Histogram Descriptor (EHD) is one of the texture descriptors of MPEG-7 that captures the spatial distribution of edges in an image. The EHD basically represents the frequency of occurrence of five different types of edges present in an image. In fact, it detects nondirectional edges as well as four directional edges. Hence, the local EHD can be a good texture signature of an image to image matching even when the underlying texture is not homogeneous (MANJUNATH; OHM; VASUDEVAN; YAMADA, 2001). Edge histograms for an image are constructed by combining sub-images into three different types of histograms: Local, Semi-Global and Global edge histograms.

The local-edge histograms are constructed from sub-images which are defined by partitioning an image into 16 non-overlapping equal units (4 x 4). According to the MPEG-7 standard, edge distribution of a sub-image is represented by a histogram using five types of edges: vertical, horizontal, 45 degree diagonal, 135 degree diagonal and non-directional. To extract these five edges, each sub-image is further divided into non-overlapping square blocks of 2 x 2 pixels. In each of these blocks, five edge oriented detectors (filters) are applied to compute the edge strength (Shown Figure 3).

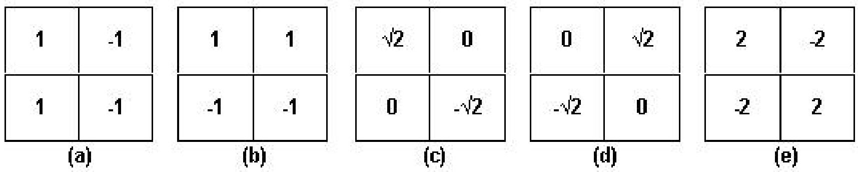
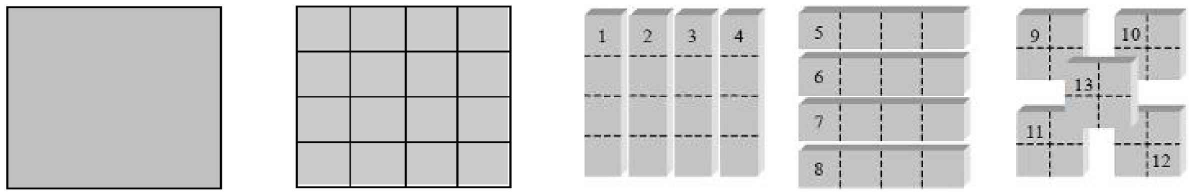


Figure 3: Five filters for edge detection (a) vertical (b) horizontal (c) 45-degree diagonal (d) 135-degree diagonal (e) non-directional edge

Thus, the edge strength of each image block is calculated to determine the type of edge. Whether an image block can be considered as an edge and its type is defined by Equation 20. Otherwise, the image-block is not counted as edge block. In other words, the image-block does not have enough strength to belong any one of the edges. If five edge strengths are defined as ES_v, ES_h, ES_{45d}, ES_{135d} and eds_{non-ed}, then the type of edge block of pixels (i, j) can be determined using the following equation (20):

$$max(ES_v, ES_h, ES_{45d}, ES_{135d}, ES_{non-ed}) > T_{edge} \tag{20}$$

In summary, using a predetermined threshold, each block is classified as an edge block and finally, its corresponding edge type. Repeating this block-based edge extraction scheme, histograms of all sub-images (16) are constructed. Hence, for each sub-image, 5 bins of local-edge histogram is generated. In order to construct the global edge histogram, the distribution of all five edges from the entire image is accumulated. Likewise, to construct the semi-global histograms, sub-images are grouped into 13 different sub-sets of image-blocks as shown in Figure 4 and its corresponding edge histograms are constructed using the local-edge histograms obtained already from 16 sub-images.



(a) (b) (c)

Figure 4: Face image partitions scheme for feature extraction (a) global image (b) Sub-images for local edge histograms (b) Subsets of sub-images for semi-global edge histograms

Combining all histograms together, the EHD of an image is represented by 150 bins (16 x 5 bins of local histograms + 5 bins of global histograms and 13 x 5 bins of semi-global histograms) which can be used to evaluate the similarity between images. The distance, D, between EHD of two images can be calculated using the Equation 22 (MANJUNATH; OHM; VASUDEVAN; YAMADA, 2001).

$$\begin{aligned} D(Im_1, Im_2) &= \sum_{n=0}^{79} |LH_1(n) - LH_2(n)| \\ &+ 5 \times \sum_{n=0}^{4} |GH_1(n) - GH_2(n)| \\ &+ \sum_{n=0}^{64} |SGH_1(n) - SGH_2(n)| \end{aligned} \tag{21}$$

The first term of equation 22 refers the local histograms where LH_1 and LH_2 represents the normalized local histograms bin values of $image_1$ and $image_2$ respectively. Similarly, the second and third terms represent the global and semi-global histograms respectively. The global term is calculated with a weighting factor 5 to increase the influence of global features between local and semi-global features (MANJUNATH; OHM; VASUDEVAN; YAMADA, 2001).

2.3 ARTIFICIAL BEE COLONY ALGORITHM (ABC)

Foraging for rich food sources to obtain maximum amount of food is the main objective of the bee colony. Honey bees have the ability to find the best food sources even in places very far from the hive. They are able to memorize the location of the food and to return back to the hive without losing their way. Bees always look for the most profitable sources among those available and can adjust their search behavior according to the changes in the environment (BAHAMISH; ABDULLAH; SALAM, 2008). The process of random search for food sources begins with scout bees (or onlooker bee) without any guidance from other bees or previous information. Whenever a scout bee finds a food source, it becomes an employed bee. The employed bees bring nectar to the hive and share the information about the food sources with other onlooker bees by means of a waggle dance. Onlookers always wait at hive to evaluate the potential of food sources so that they can select the best one to collect food (TERESHKO; LOENGAROV, 2005). The exchange of information among bees to form the collective knowledge takes place only in the dancing area. Onlookers probably may watch many dances and may choose to employ itself at the most profitable source. Therefore, the rate of recruitment which defines the quantity of bees that should participate in the nectar collection varies according to the profitability of a food source (TERESHKO; LOENGAROV, 2005).

The overall behavior and efficiency of a bee colony to strive in a complex, challenging and changing environment has inspired a computational approach for optimization. Motivated by this foraging behavior of honeybees, (KARABOGA, 2005) proposed the artificial bee colony algorithm. In the next section, this population-based optimization algorithm is briefly detailed.

2.3.1 ABC ALGORITHM

In the ABC algorithm (KARABOGA; AKAY, 2009), each food source is considered as a possible solution for an optimization problem. The nectar amount represents the quality (fitness) of the solution represented by a food source. At the beginning, the number of employed bees and onlooker bees must be defined, usually the same. The quantity of employed bees represents the number of the solutions (SN) in the population. The algorithm starts by associating all employed bees to randomly generated food source positions that are considered as an initial population of SN. Each solution is represented by X_i, such that $i \in (1,2....SN)$. Each X_i is a d-dimensional vector, and d represents the number parameters to be optimized. Once the employed bees are created, the search process starts and it is repeated by a predefined number of cycles, represented by MCN (Maximum Cycle Number).

During the search, an employed bee looks for a new food source modifying its position in the search space, and then evaluating the amount of nectar at the new source. If the new amount is higher than the previous one, then the position of the old source is forgotten and the new position is memorized. When the employed bees complete the search, they will share their food source information with the onlooker bees, and the process will be repeated by the onlookers. The algorithm consists of three main steps in each cycle or generation:

1. Sending the employed bees to look for food sources and evaluating their amount of nectar;
2. Sharing, by dance, the food source information of the employed bees with onlookers which, in turn, select the food sources and evaluate their nectar amount;
3. Determining the scout bees and sending them randomly to search for new food sources.

Whenever a bee, whether it is scout or onlooker bee, finds a new food source, then it becomes an employed one. Whenever a food source is fully exploited and exhausted, all the employed bees associated with it abandon the food source and will become scouts or onlookers.

As previously explained, the ABC is an iterative algorithm. Then, during each cycle or generation, every employed bee moves to the neighborhood of its currently associated position and evaluates its nectar amount. If the nectar amount of its neighborhood is better than that of its current position, the employed bee leaves the current position and moves to the new position, i.e., memorize the new position and forgets the old position. Otherwise, it just stays in its old current source position. When all employed bees finished their neighborhood search, they fly back to the hive to share their information about the amount of nectar of the food sources found recently with onlookers that are waiting in the dance area. Each onlooker selects a food source position according to a probability related to its nectar amount.

After all onlookers have selected their food sources, each of them determines a food source in the neighborhood of its current position and evaluates its fitness. Again, like employed bees, a greedy selection mechanism is employed to select the new food source between neighborhood and current position. If a solution cannot be improved during a predetermined number of cycles, it is abandoned by its employed bee, which becomes a scout. Then, this scout bee will go for searching a new food source position randomly and the new random solution will replace the abandoned one. The selection of a food position that should be abandoned is determined by a *Limit* variable associated with each solution. The *Limit* is defined by $d \times SN$. The whole process is repeated for a predetermined MCN or until a termination criterion is reached.

In this model, scout bees can be seen as performing the job of exploration or global search, whereas employed and onlooker bees can be seen as performing the job of exploitation or local search (KARABOGA; AKAY, 2009). The basic ABC has three control parameters whose values should be determined at the beginning of the search process: *SN*, *Limit* and *MCN*.

2.3.2 ABC ALGORITHM FOR THE FACE DETECTION AND RECOGNITION PROBLEM

An object image is represented by a 4-tuple (x, y, s, θ) (coordinates x and y, scale and rotation angle) as defined in section 2. These four transformation parameters should be optimized to find out the most similar object image in a still cluttered image with multiple face images. By considering our image context, the search space is limited by restricting the range of the parameters as follows: $x = [0..n]$ (column), $y = [0..m]$ (row), $s = [0.5..1.5]$ (scale), $\theta = [-\pi/2..\pi/2]$ (rotation).

A bee or solution is a set of (x, y, s, θ) representing a position in the still image. Then, each solution can be represented by a 4-dimensional vector $X_{id}(X_{i1}, X_{i2}, X_{i3}, X_{i4})$, where $d = 4$ and $i \in 1, 2, 3....SN$. A new position (food source) in the neighborhood of a specific solution is determined by altering the value of one randomly chosen solution parameter of X_{id} and keeping unchanged the remainder parameters. This neighborhood position can be calculated by adding to the current value of the randomly selected parameter, the product of a random number between $[-1..1](\theta)$ and the difference in values of this parameter position and some other randomly chosen position of the same parameter X_k. Variable k is determined randomly between $[1..SN]$ and should be different from i. In order to determine a solution X_i' in the neighborhood of X_i, a solution parameter j and another solution $X_{kj} = (X_{1j}, X_{2j}, X_{3j}, ...X_{kj})$ are selected randomly. Except for the value of the selected parameter j in range $[1..d]$, all other parameter values of X_i are same. The value of the new neighborhood position X_i' is calculated using the following Equation (23):

$$X_{ij} = X_{ij} + \phi(X_{ij} - X_{kj}) \tag{23}$$

where k, $i \in (1..SN)$, ϕ is a random number in the range $[-1..1]$ and k is a random index that should be different from i. The probability p_i of selecting a food source i by an onlooker can be calculated by Equation (24):

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \tag{24}$$

where fit_i is the fitness of a solution. After the abandoned solutions are determined using *Limit*, the new random scout bees can be produced using the following Equation 25:

$$X_{ij} = Xmin_j + rand(0,1)(Xmax_j - Xmin_j) \tag{25}$$

where $Xmin$ and $Xmax$ represents the lower and upper bound values allowed for the four parameters (x, y, s, θ) and *rand* is a random value in the range [0..1].

The following pseudocode summarizes the ABC algorithm implemented in this work:

1. Initialize the population of solutions of employed bees (positions) X_{id} of size *SN*;
2. Evaluate the fitness of each element of the population by calculating the similarity measure by TM of the RI and target cut from still image;
3. Repeat until the stopping criteria is met (cycle = *MCN*):
   (a) Produce new neighborhood solutions of image positions X_i' for each X_i of employed bee using (23) and evaluate their fitness;
   (b) Move the employed bees to new positions (solutions) by selecting the best among neighborhood positions and current positions respectively;
   (c) Calculate the probability values p_i for the solutions X_i, using (24);
   (d) Produce the new solutions X_i' for the onlookers using the probability values p_i from the solutions X_i and compute their fitness;
   (e) Move the onlooker bees to new positions applying the selection process to the neighborhood positions and current positions respectively;
   (f) Abandon the stagnated solutions, transform the corresponding bees to scouts and replace them with new randomly produced solutions using (25);
   (g) Increment the cycle;

During the steps 3a and 3d, employed and onlookers bees carry out the local search process for optimal solutions, meanwhile, in step 3f, scout bees are generated to perform global search in the whole search space aiming at finding new unexplored solutions. For each solution, there is a limit counter value to check whether it improves during the generations or not. If it does not improves, the counter is incremented in each generation. At end of the each generation, when a counter value of a specific solution reaches the textitLimit, then it will be selected to be abandoned and substituted by a scout bee. If none of the solution reaches this *Limit* value, then no scout bee will be generated.

2.3.3 IMPROVED ABC ALGORITHM (IABC)

According to the the basic ABC algorithm, only one parameter is perturbated at a time when a new neighborhood solution is generated for both employed and onlooker bees. In the scout bees phase, only one scout bee is generated when the limit counter value of a specific solution exceeds the *Limit* parameter. Otherwise, no scout bee will be produced. During the optimization process, the population of solutions can converge to a sub-optimal region in search space and resulting in stagnation of the best solution for a certain number of cycles continuously. When occurs stagnation, to restart the search, by keeping always the best solution, explosion procedure is applied in the object recognition problem (PERLIN; LOPES; CENTENO, 2008). The explosion procedure generally aids the algorithm to search in different region of search space and to find the best solution gradually during the iterations.

Our main objective in the FR study is to recognize a face object image as fast as possible so that this type of algorithm can be applied to real-world problems. Based on this context, therefore, to define the improved ABC algorithm, three main mechanisms were tested in the study conducted by Chidambaram and Lopes (CHIDAMBARAM; LOPES, 2010): (1) the perturbation of all variables; (2) generation of scout bees; (3) explosion of stagnated solutions. Based on the three proposed new mechanisms and the mechanisms which are already present in the basic ABC algorithm, such as the perturbation of one variable and generation of one scout bee through the Limit parameter, several experiments were done. Combination of these mechanisms resulted in eight different strategies. Finally, among the all strategies, the best strategy was determined by evaluating the results of all experiments from the object recognition problem developed by Chidambaram and Lopes (CHIDAMBARAM; LOPES, 2010). The best strategy consists of perturbation of all four variables, without generation of scout bees and with the explosion of stagnated population. Hence, in this work, we have used the improved ABC algorithm to recognize the faces in still images.

2.4 RELATED WORKS

Many advances have been made in recent years, achieving FR rates higher than 90%. Nevertheless, the face image acquisition process undergoes a wide range of variations due to the development of new technologies. Comprehensive studies of FR techniques have been done by Chellappa (CHELLAPPA; WILSON; SIROHEY, 1995), Zhao (ZHAO; CHELLAPPA; PHILIPS; ROSENFELD, 2003), Tolba (TOLBA; EL-BAZ; EL-HARBY, 2006) and Abate (ABATE; NAPPI; RICCIO; SABATINO, 2007) with their colleagues.

Predefined templates (represented by objects such as eyes, nose or the whole face) that represent the features of target faces are used to find similar images (GUO; YU; JIA, 2010). Since the eyes are one of the most relevant components when dealing with frontal face features, their previous detection may reduce the computational cost (CAMPADELLI; LANZAROTTI; LIPORI, 2008; GUO; YU; JIA, 2010). Geometrical measures of the face configuration were used in many works to detect interest points. These features and the related distances can be used to build a robust facial structure that is suitable for FR in large image databases. The template matching and geometric-based features methods are widely applied to FR applications for many years. However, the geometric-based features method can be easily affected by the measurement process. Likewise, the template matching based methods are highly sensitive to the environment, size and pose. Similarly, many other appearance-based methods (PCA, LDA, etc.) have been developed during the past two decades (TAN; CHEN; ZHOU; ZHANG, 2006). Among these, PCA-Eigenfaces was well-known technique used for FR. Many of the appearance-based methods are also affected by variations in lighting conditions, pose and expressions (BELHUMEUR; HESPANHA; KRIEGMAN, 1997). To minimize these variations, an analytical appearance based method named Elastic Bunch Graph Matching (EBGM), was proposed by (WISKOTT; FELLOUS; N.KUIGER; MALSBURG, 1997). EBGM compensates the non-linear characteristics not addressed by LDA-Fisherfaces and PCA-Eigenfaces. Although this method is considered as one of the most successful techniques, it is a computationally intensive algorithm, because of the exhaustive search stages spanning the entire image.

Besides, many attempts were done to attack the illumination variation problem. Variation of illumination is one of the main issues of FR. Many approaches have been proposed to handle this issue such as applying intensity normalization procedure (BRUNELLI; POGGIO, 1993), Local Binary Pattern (LBP) (OJALA; PIETIKAINEN; HARWOOD, 1996), preprocessing algorithms (GROSS; BRAJOVIC, 2003) and plane subtraction with histogram equalization (SOLAR; QUINTEROS, 2008). Some other works were focused specifically on FR of single face images using image-filtering techniques (ARANDJELOVIC; CIPOLLA, 2009; NABATCHIAN; ABDEL-RAHEEM; AHMADI, 2011) and on holistic approaches (MATSUMOTO; SHIRAI; SHIMADA; SAKIYAMA; MIURA, 2006). Most of the algorithms studied are based on the relative values between neighbor pixels and/or the local mean intensity of pixels. Although they have used specific algorithms, in fact, the illumination compensation and normalization with local information improves the recognition performance. In addition to these techniques, the basic images features such as color and edges are also used in FR.

In the literature, there is a large number of publications using edges, because its extraction is similar to the human visual recognition process. Karungaru et al. (KARUNGARU;

FUKUMI; AKAMATSU, 2004) uses the edges as the most relevant features. A T-shaped edge template covers the eyes, nose and lips by using the Laplacian of Gaussian and 3 x 3 mean filters. Although this kind of procedure is invariant to small variations in illuminations, the feature extraction may be easily affected by scale, rotation and pose. Using the MPEG-7 standard, (RAHMAN; NAIM; FAROOQ; ISLAM, 2010) applied EHD to FR and classified the images using PCA. They tested different sizes of partitioned images and found that the local features are most useful than semi-global and global features and horizontal division of images provide the best performance.

In the FR literature, color has been used to locate face regions, for example. A skin color information can be useful to differentiate a face region from complex backgrounds which can probably reduce the computational effort spent in complex background regions (LIN, 2007). (HSU; ABDEL-MOTTALEB; JAIN, 2002) proposed a face detection algorithm that is suitable for static color images using a lighting compensation technique and a nonlinear color transformation in the YCbCr color space. (LIU; LIU, 2008) proposed a novel hybrid Color and Frequency Features (CEF) method for face recognition, which derives the complementary features in the frequency domain of the component images in the hybrid color space RIQ. This color space is the combination of the R component of RGB, and the chromatic component of I and Q of the YIQ2 color space. It is based on the criteria that the use of complementary features of color spaces may improve the face recognition performance, and fusing features across color spaces can also enhance the discrimination power (LIU; LIU, 2008). Though the color and edges are the most commonly used features in many Computer Vision applications, nowadays, image features represented by interest point descriptors are frequently used in FR and related tasks.

Instead of traditional object recognition tasks, face processing tasks using interest point detectors seems to be a promising area. (FERNANDEZ; VICENTE, 2008) have applied the Harris-Laplace detector (MIKOLAJCZYK; SCHMID, 2004) and the Difference of Gaussian detector (LOWE, 2004) to the face recognition problem. They suggested that these detectors produce better results than appearance-based approaches. Another work developed by (ASBACH; P.HOSTEN; UNGER, 2008) provide a qualitative and quantitative analysis of some interest point detectors such as Hessian, Harris (MIKOLAJCZYK; SCHMID, 2004), Difference of Gaussian (DoG) (LOWE, 2004) and Laplacian of Gaussian (LoG) (MIKOLAJCZYK; SCHMID, 2004), SURF (BAY; ESS; TUYTELAARS; GOOL, 2008), and SIFT (LOWE, 2004) in face detection and localization. According to the results, the Harris interest point detector with SURF descriptors was considered as the most promising combination. In a recent work for iris recognition, features are extracted using SURF to handle issues such as partial occlusion, non-uniform illumination and head tilt during acquisition (MEHROTRA; SA; MAJHI, 2012).

Depending on the application, the localization feature points may need some iterative search algorithms (ETEMAD; CHELLAPPA, 1997). In general, swarm intelligence algorithms have drawn the attention of image segmentation, image retrieval, object recognition and face recognition research communities. They have been used for feature extraction and feature optimization. In addition to the classical swarm intelligence algorithms such as ACO (BONABEAU; DORIGO; G.THERAULAZ, 1999) and PSO (KENNEDY; EBERHART, 1995), ABC (KARABOGA, 2005) optimization has also been successfully applied to solve many optimization problems, for example, (CHIDAMBARAM; LOPES, 2010).

One of the application of swarm intelligence algorithms in FR is the the feature selection from the face images. Feature selection process can be initialized heuristically with a subset of features and more features are added iteratively. (RAMADAN; ABDEL-KADER, 2009) have used PSO to optimize features selection in a face recognition system. Feature selection in pattern recognition can be used as a pre-processing step in order to reduce the irrelevant features before classification. This same problem was also reformulated into an ACO approach by (KANAN; FAEZ; HOSSEINZADEH, 2007). A similar work was proposed for face recognition using both GA (Genetic Algorithms) and ACO in which the ACO is used to extract features, and the recognition is done by GA (VENKATESAN; MADANE, 2010). Another approach using hybrid Taguchi PSO can be found (LIN; CHU; LEE; HUANG, 2008) in which face recognition was done using neural networks. In this work, with the objective of gathering information from images in both the space and frequency domains, they extracted local features using Gabor wavelets at different scales and orientations.

Nowadays many FR works select the features from local and global regions of images or, more generally, from sub-images. Generally, the size of a sub-image can determine the boundary between local and global regions. The increase in size of local region may increase the globality of local features. Currently, many FR works have been focused on global and local features based methods. Most of the subspace methods (PCA and LDA, for example) based on dimensionality reduction fall into the category of global feature based methods. However, these methods are very sensitive to the global changes of images, such as illumination variation and expression (ZHOU; AHRARY; KAMATA, 2012). In order to overcome these issues, many matching methods using local features are widely developed in FR applications. One of the well-known methods is LBP (Linear Binary Pattern)(OJALA; PIETIKAINEN; HARWOOD, 1996). LBP is fully exploited and adopted in numerous FR applications. Using global and local features, (AMARAL; THOMAZ, 2012) have proposed an FR approach comparing LBP (Linear

Binary Pattern) and PCA (Principal Component Analysis) in order to understand the behavior of both methods when using pre-processed and previously spatially normalized face images. In another work, (AHONEN; HADID; PIETIKAINEN, 2006) described the face using LBP. The idea behind this descriptor is that the face can be seen as a highly discriminative micropattern. However, the micropatterns derived from the local neighborhood pixels encode only one kind of spatial information around the local pixels.

Another attempt to overcome the issues related to images was proposed by (ZHOU; YANG; PENG; WANG, 2006) representing the faces through holistic as well as local information. Holistic features were extracted from the whole face, meanwhile local features were extracted from sub-images using Discrete Cosine Transform (DCT). Using the improved LDA (Fisherfaces), the training and classification was done using the ORL and Yale face databases. In this work, 96.9% of FR accuracy was obtained. Similarly, a robust two-stage FR using global and local features was developed by (SINGH; WALIA; MITTAL, 2012). Global features were obtained from Zernike moments (radial moments) method and local features were extracted from the histogram-based Weber Law Descriptor. Zernike moments are moment invariants descriptors used in character recognition, for example. The other Weber Law Descriptor was obtained by incorporating LBP representation along with relevant information from edges (SINGH; WALIA; MITTAL, 2012). Using the ORL and Yale databases, they obtained 97.85% and 92.11% recognition rates, respectively.

In order to boost the FR performance using complementary facial information, (LIU; LIU, 2010) proposed a FR approach fusing multiple face features derived from the hybrid color space. From this space, multiple features were derived using three image encoding methods such as Gabor representation, LBP and DCT (Discrete Cosine Transform) of the input image. Complementary information was obtained by fusing color, local and global frequency information. The proposed method achieves a face recognition rate of 92.43%. In addition to this work, (GENG; JIANG, 2011) proposed a framework of FR based on the multi-scale local structures of the face image using SIFT. The main motivation of this work was based on the fact that the multi-scale local features have the power to be more robust against image variations and other related issues than the holistic approaches. Besides, a single approach is not expected to deal with complex FR problems. Rather, well-designed individual methods are essential to obtain high performance rates. In the same work, (GENG; JIANG, 2011) proposed a two-stage image matching scheme and a strategy of keypoint search for the nearest subject. Finally, FR is performed through a training procedure for multiple samples per subject. According to the results, the recognition rate of the proposed framework overcomes the SIFT achieving above 94% meanwhile SIFT generates above 80%.

Similarly to the previous works, (WANG; YANG, 2008) proposed an approach in which they have implemented a simplified hierarchical face detection method by using Template Matching and 2DPCA algorithm (YANG; ZHANG; FRANGI; YANG, 2004). A significant improvement in the recognition performance was achieved using this approach in images with many faces. In order to improve the generalization capability of the LDA, (KIM; KIM; HWANG; KITTLER, 2005) developed a component-based LDA representation in which a face image was partitioned into forehead, eyes, nose and mouth regions. In their proposal, a cascade LDA scheme is demonstrated which combines the component-based LDA and the holistic LDA applied to a whole face at the feature level.

As shown in the section, it can be noticed that the feature extraction methods such as SURF and EHD were already used in previous works, except CA. Likewise, swarm intelligence algorithms such as ACO and PSO were applied in the applications of FR, mainly for feature selection. Though ABC was used in application like object recognition, in the context of the MFR, no work can be found in the literature. To the best of our knowledge and, according to the previous works, FR approaches are mostly based on supervised approaches with some sort of training schemes.

3 METHODOLOGY

In the present chapter, two main approaches of FR are presented : (1) the novel semi-supervised approach for MFR; (2) the hierarchical approach for SFR. The main aspects of these two approaches are explained in the following sections.

3.1 MULTIPLE FACES RECOGNITION APPROACH (MFR)

The main focus of the MFR approach is to deal with still images having multiple faces, where each interest region may be subject to different illumination conditions and the background is typically complex. Although there are several stages to recognize faces in this kind of images, such process can be summarized in two main steps: (1) the searching for interest regions in the still image, from which the optimal local features can be extracted; (2) face matching and recognition based on the extracted features. These two steps can be seen in the general view of the proposed approach as shown in Figure 5. At each iteration, the searching process defines a different interest region on the still image. The recognition stage estimates the similarity between the features extracted both from the original face object image (query) and the interest region. The recognized image corresponds to the most similar region after a given number of iterations.

As shown in Figure 5, the searching stage in which the interest regions are searched on still images. This stage is similar to the template matching procedure (GONZALEZ; WOODS, 2009), as shown in Figure 6. However, in this stage, instead of matching pixel to pixel values as template matching, the features extracted from interest regions and face object image are matched. Interest regions with central coordinates (x, y) and having the same size as the template (face object image) are submitted to a matching process. The dotted lines indicate the limit for the center of the candidate interest regions so that they do not exceed the image border. Thus, the padding is limited to half of the template *width* (w) and of the template *height* (h) - in the horizontal and vertical borders, respectively . Hence, the search space is limited to the Min(x, y) and Max(x, y) coordinates.

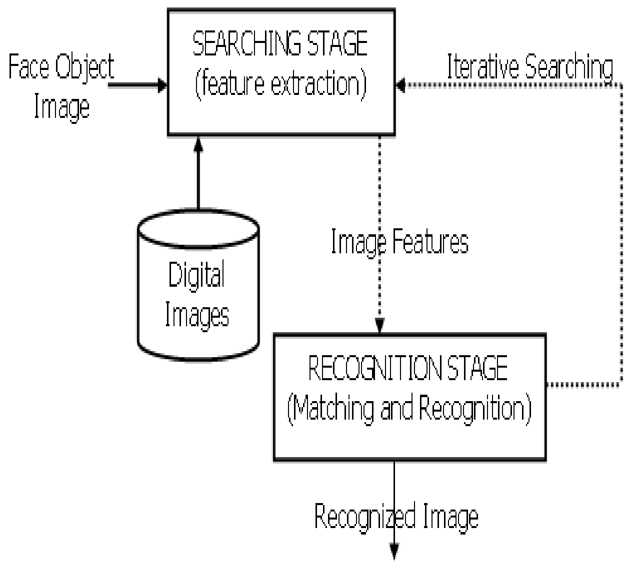


Figure 5: General view of proposed FR approach

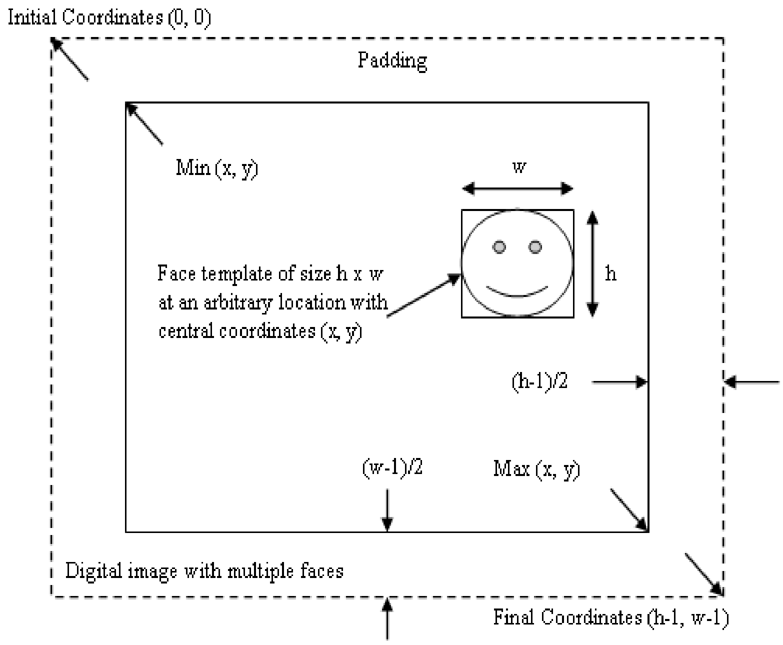


Figure 6: FR scheme using template

In this work, the matching process is based on the swarm intelligence algorithm named iABC, similar to template matching which was already applied for object recognition and proved to be an excellent optimization algorithm (CHIDAMBARAM; LOPES, 2010). Each individual is represented by a 4-tuple representing planar coordinates (x and y), rotation angle (θ) and scale factor (s).

The search process follows the method described in Section 2.3. Initially, two parameters must be set: the initial population (SN) and the maximum cycle (or generation) number (MCN). SN represents the number of individuals of the bee population and MCN determines the number of iterations of the optimization process. Besides, the number of times the iterative process should be repeated is defined by the parameter number of runs (NRUNS).

During the iterative process, each and every target image (interest region) is cut using the parameters of individuals of the population. The iABC algorithm will try to find the face using the maximum similarity among all individuals of the initial population and consequently, the optimal parameters related to each face object image can be determined through the search process.

The iABC algorithm also needs the definition of the stagnation (or convergence) and decimation factors (discussed in Section 2.3). During the iterative process, stagnation of individuals affecting the search for optimal solutions (best match for face object image in still images) may happen. A solution to the stagnation condition can be done using explosion (decimation) mechanism. The decimation is related to generation of new individuals and substitution of some part of the population.

As already mentioned, the matching stage evaluates the similarly between the features extracted both from face object image and each interest region. In this work, the feature extraction stage considers SURF, a detector and descriptor invariant against changes in scale, rotation and brightness. This method is robust and efficient, and has shown to be significantly better than other similar approaches, for example, SIFT (PIMENOV, 2009; BAY; ESS; TUYTELAARS; GOOL, 2008).

To evaluate the similarity between two face images based on the SURF descriptor, the repeatability rate will be used, a measure of stability that is strongly accepted as a standard computer vision performance metric for interest points (TRUJILLO; OLAGUE, 2006). In summary, the percentage of points repeated in the two images being compared is defined as the repeatability rate. It is pertinent to recall that a point is considered repeated if it lies in the same location (around the same coordinates) on both images.

Since considering an exact matching between the interest points' coordinates is not feasible, an acceptable error margin must be determined. In fact, two thresholds need to be defined: one related to the coordinate distance error (T_{CDE}) and other to the descriptor distance error (T_{DDE}) - as discussed in Section 2.2.1.3.

The repeatability rate corresponds to the fitness, a well-known denomination in optimization approaches that will be considered in the following text. Hence, the similarity between the face object image and each interest region is represented by the fitness.

In summary, the entire recognition process is based on two main steps under an iterative procedure. For each cycle until a predetermined number of runs (NRUNS): (1) apply iABC to define interest regions in the still image for feature extraction (interest points and descriptors) using SURF and (2) determine the fitness based on the similarity between the face object image and the interest regions based on the repeatability rate.

In each cycle, the individual with the maximum fitness among all population will be selected as the best individual. Finally, considering the population of best individuals, the individual with maximum fitness will be selected as the best solution (corresponding to the face image with maximum similarity).

This process can be seen as an optimization problem involving several steps of window-based processing and matching. During the iterative process, each cut and comparison of image may lead to find the face object image. Since the parameters involved here may take a large range of values, the number of possible combinations may become very large. In addition, the FR stage may become complex due to the parameters involved in the other feature extraction methods. The search space will certainly increase according to the number of possible solutions tested and the size of the still image. Consequently, the computational cost will increase according to the number of comparisons. Hence, this approach is treated as an optimization problem.

Observe that the FR process using SURF-iABC approach requires many parameters, namely stagnation, decimation, T_{CDE} and T_{DDE}. Hence, the definition of the parameters becomes part of the methodology. The main motivation for the definition of thresholds is to determine the values that can yield the highest recognition performance. Initially both parameters should be defined through a threshold analysis experiment so that they can be set at the beginning stage of all FR experiments. In fact, the threshold analysis is done on a set of still images from different conditions and its recognition rates are evaluated. From the results, the values that yield the highest recognition rate will be selected for the further FR experiments.

In the described MFR approach, the interest regions are determined in each cycle and the corresponding features are extracted and compared to that representing the face object image. However, to improve the computational time, one can determine the SURF features and descriptors of the entire still image at the beginning of FR process. In this scheme, the interest points of the entire still image can be transferred to a separate matrix structure with the same size of the still image. This approach is denominated as Matrix SURF-iABC. During the matching process, the interest points for each interest region can be obtained from the static matrix structure at the same coordinates. However, the effective functioning of the scheme should be evaluated to check the recognition performance for varying image conditions. Furthermore, both SURF-iABC and Matrix SURF-iABC approaches should be compared so that at what image conditions they effectively generate higher recognition rates.

Although there are many ways for searching and recognizing faces using SURF in still images, in this work, it is combined with the iABC algorithm. In the next chapters, the proposed approach will be discussed showing how it can aid effectively and efficiently the FR process on images having multiple faces and acquired under varying conditions related to illumination, pose and expression. Note that other feature extraction methods can be considered other than SURF, including CA and EHD. In addition to the use of SURF and iABC, another different aspect of this work is that this is a semi-supervised learning approach entirely based on the discriminative power of local features obtained from interest points.

3.2 SINGLE FACE RECOGNITION APPROACH (SFR)

The SFR process solves the problem of identification of a given unlabeled face image which is compared to images from a database of known individuals. In other words, the main task is to query a face image obtained under different image conditions into a database of face images obtained under controlled condition, for example in a criminal investigation scenario. In the present approach, recognition will be done comparing individually all extracted features of an query image with features of the database. The image conditions of query images referred to different illumination conditions and with variations on expression and scale.

The main task of fully automatic FR systems is shown in Figure 7 which initially consists of image acquisition and normalization, and then followed by feature extraction and recognition. Image acquisition and normalization stages are done in previous works (PRODOSSIMO; CHIDAMBARAM; LOPES, 2012, 2013) in which normalization is done through eye detection. In the present work, the first main task concerns the extraction of facial features. Extracted face features should confirm the similarity between base images and query images with a minimum error. When a query image is presented, to find out a similar one, an appropriate feature representation and a similarity measure to rank the images is necessary (DATTA; JOSHI; LI; WANG, 2008).

Facial features can be affected by image variations and consequently, they may lead to errors and low accuracy. Since feature extraction methods should provide a sufficient and meaningful set of features to be used in the FR tasks, the dependence on the image variations should be minimized. Faces typically have the same facial components but with some specific variations inherent to each individual which includes skin color, shape variations of facial

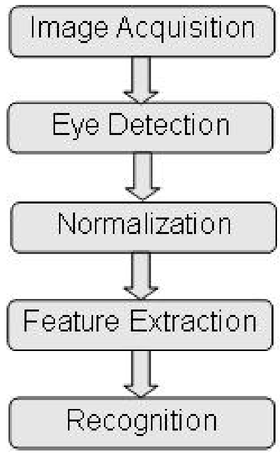


Figure 7: Main tasks of automatic FR system

components, for example. Geometrical measures of facial components are subject to errors due to image variations. It is also hard to measure the geometrical shapes and specific features of facial components. Therefore, in this SFR approach, the faces are treated generally as images without any consideration on facial components such as eyes, mouth, etc. Features will be extracted using three feature-based methods: SURF, EHD and CA. More details regarding these methods can be seen in Section 2.2. The SFR process and its sub-tasks are shown in Figure 8. in which, the query image features are matched against all feature vectors of database images. The final decision is done using a classification algorithm. Based on this procedure, the present SFR approach is developed.

Using EHD and SURF, the features are extracted following the methodology presented in Section 2.2. SURF naturally extracts only local features from interest points. From EHD, three kind of features can be extracted according to the partitioning of the image: global, semi-global and local features. The CA method is originally based only on global features,which have only the potential of generalizing an entire image. The local features can preserve some spatial information and are considered as robust against face expressions, noise, and occlusion (SINGH; WALIA; MITTAL, 2012). Therefore, the same image partitioning scheme of EHD is also applied to the feature extraction procedure of CA. Then, the color angles will be extracted and combined in one descriptor vector with 90 bins which corresponds to 3 global, (16×3) local and (13×3) semi-global angles (shown in Figure 4). Based on the methodology of features extraction exposed in this section, the present SFR is proposed. The schematic diagram of SFR approach is shown in Figure 9.

Features that are identified by the SURF detector as distinctive and stable are normally stored in a descriptor vector of 64 elements for each interest point. The number of interest points may vary due to the type of face image and its variations. Consequently, the size of a

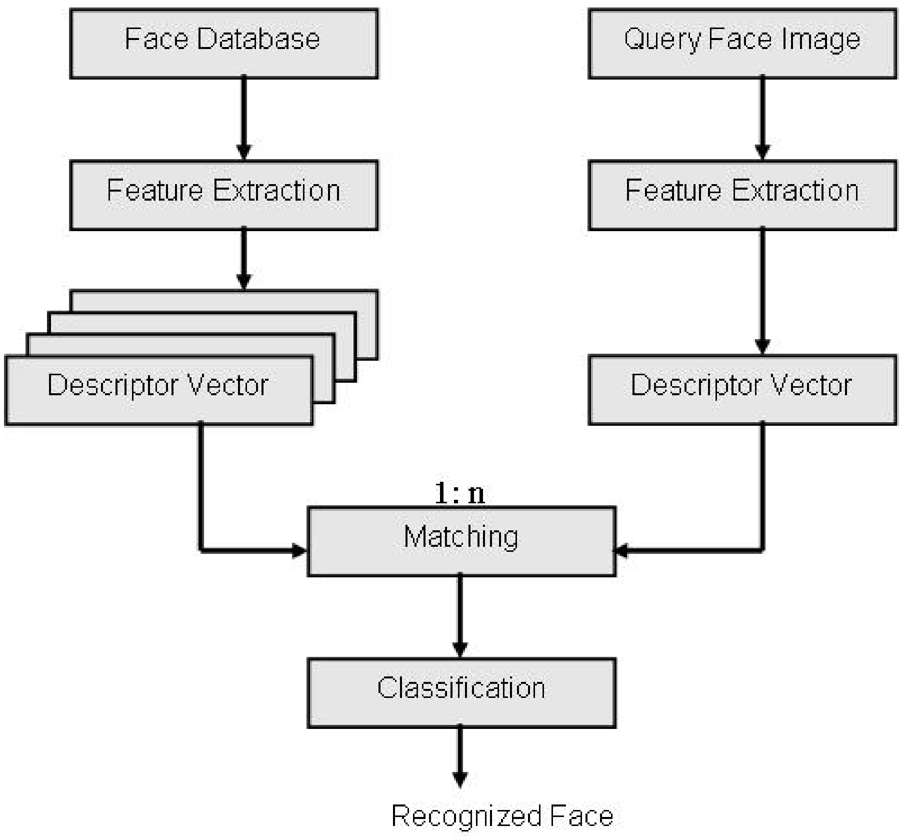


Figure 8: Schematic diagram of SFR approach

descriptors vector of a face image certainly depends on the number of interest points. On the other side, the size of EHD and CA is always of 150 elements. Applying the proposed features extraction scheme, first, the database of descriptors of base face images will be constructed.

Unlike the SURF method, the SFR approach using CA and EHD is analyzed by different combinations of G (G), SG (SG) and L (L) features, namely, G+SG+L, G, SG, L, G+SG and SG+L. The main objective of the feature combinations is to evaluate the influence of different features in recognition performance. In this work, the SFR approach will be evaluated in two ways: (1) Independent analysis of CA, EHD and SURF methods; (2) Hierarchical approach with two sequences of combinations using all methods.

The CA, EHD and SURF descriptors are first considered independently, in order to evaluate and to fully exploit the potential of the complementary information provided by each method. In order to quantify the discriminative capacity of methods individually on FR, the single-stage analysis becomes important. Both the recognition rate and computational effort will be considered during the evaluation of SFR approach.

Besides robustness and distinctiveness, FR methods must also provide a fast processing and retrieval. To reduce the computational time while taking advantage of the complementary

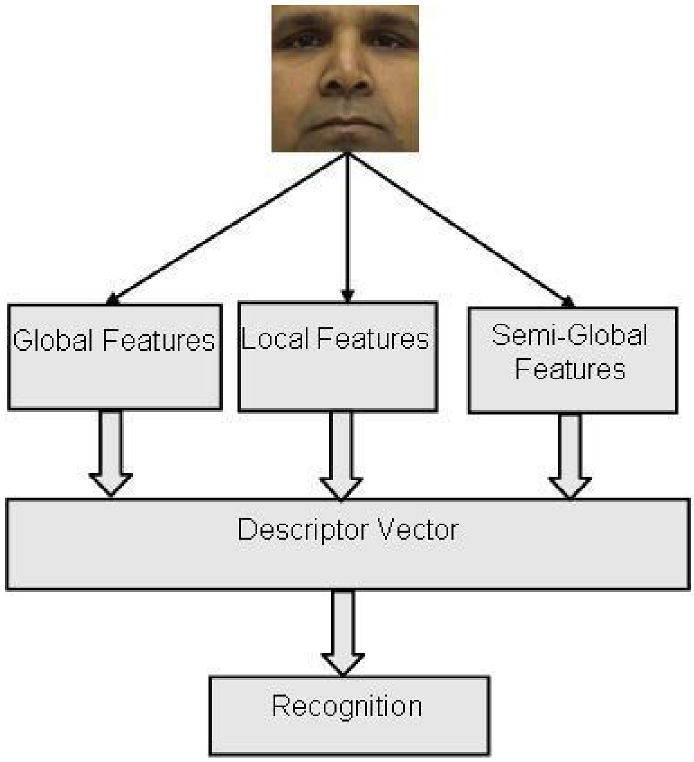


Figure 9: Schematic diagram of features combination

information provided the different descriptors, the hierarchical approach is proposed. This is the one of the main contributions of our research. In this approach, the feature extraction and matching procedures will be performed in three stages through two sequences of methods: (1) EHD, CA, and SURF; (2) CA, EHD and SURF. One of the motivation to attempt the hierarchical approach is to reduce the computational effort, therefore, the third stage will be conducted only with SURF, which is the most computationally intensive. The hierarchical SFR scheme is shown in Figure 10.

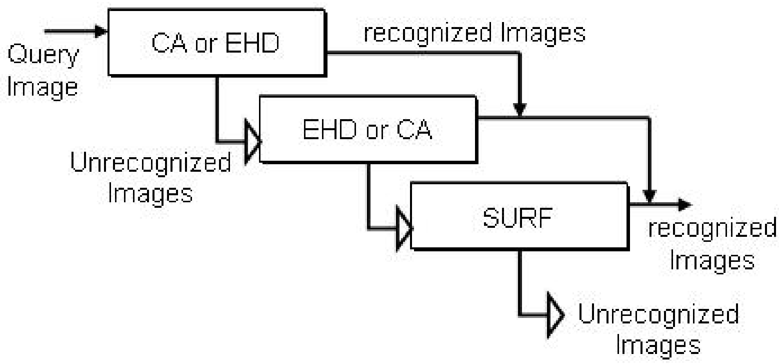


Figure 10: Proposed hierarchical SFR approach

In summary, in the hierarchical approach, only the query images with incorrect matches are processed in the subsequent stages. To evaluate the robustness of each method, query face images are acquired under different scenarios, including different lighting conditions, face expression and scale.

During the querying process, the descriptor vector of the query face image will be compared with the descriptor vector of all face images in the database (1 to n). The matching procedure is based on the Euclidean distance measure. In other words, the retrieved descriptor vector that presents a minimum distance among all comparisons will be considered as a correct match (similar to k-nearest neighbor in which k = 1). This procedure is applied to the EHD and CA. But, in the case of SURF, the selection of a correct match depends on the repeatability rate as defined Equation 3. The repeatability rate determines the maximum similarity between two face images. Unlike EHD or CA, to select the best match, the maximum repeatability rate among all comparisons should be considered. Furthermore, the Euclidean distance measure of descriptor vector of query image and database images must respect the threshold distances that are defined by Equations 1 and 2. To produce high recognition rates, two distance thresholds, coordinate distance (T_{CDE}) and descriptor distance (T_{DDE}), must be defined through preliminary experiment analysis

Similar to threshold distances of the SURF, the EHD also needs the definition of a threshold for edge strength (Equation 20) through an experimental analysis. However, the threshold of EHD is necessary only for the definition of types of edges from 4 pixels (2x2) image blocks, not for the distance measure between descriptor vectors.

In the literature, the SFR applications are mostly proposed using machine learning algorithms in combination with a training scheme on large databases. This kind of method requires certainly additional computational effort due to the learning process. In this work, the SFR is strictly based on the discriminative power of extracted features. In each stage of the hierarchical FR process, face names are required to check whether a face is correctly classified or not. Though previous knowledge is necessary, to verify results, neither training nor learning algorithms are required. Hence, this approach can be considered as a semi-supervised approach. This kind of procedure avoids unnecessary computational effort, because not all methods will be used together in the same stage neither all images will go through the three stages. Once an face image is classified, it will be removed from the base image list and will not go through other stages. Besides, to evaluate the influence of the type of features in FR performance, all possible combination of global, semi-local and local features will be tested. This proposal is based on the fact that the alternative approaches with heterogenous features and their complementary information can somehow contribute to improve the FR performance. Although several works have been proposed for recognizing single faces from databases in recent years, for the moment, considering the methodology followed and the feature extraction methods used in this approach, no similar work was not found in the literature.

4 EXPERIMENTS AND RESULTS

In this chapter, the two main approaches, MFR and SFR, are presented. In Section 4.1, the recognition of multiple faces in still images and in Section 4.2, the recognition of single faces from data base are discussed. In both sections, image preparation and experiment details, threshold analysis and experiments with images obtained under different conditions are explained.

4.1 MULTIPLE FACES RECOGNITION (MFR)

All experiments and results regarding MFR are widely discussed in the following sections. The tuning of parameters and thresholds for SURF and iABC are first discussed. Then, a preliminary experiment analysis using Matrix SURF-iABC and SURF-iABC approaches are detailed. The experiments based on all image conditions and the experiments with face object images, that evaluate the potential of the proposed approach are explained in Subsections 4.1.4, 4.1.6 and 4.1.5. Finally, a brief discussion on results is done in Subsection 4.2.5.

4.1.1 IMAGE PREPARATION AND EXPERIMENT DETAILS

Based on real-world conditions, eleven different categories of still images were prepared for the experiments reported in the following sections. Still images with multiple faces were captured under three main illumination conditions: (1) using a specific lighting system with two light sources, denominated as Illum-I (Experiment I); (2) using a specific lighting system with one light plus room lights (fluorescent lamps), denominated as Illum-II (Experiment II); (3) under room lighting conditions (fluorescent lamps) denominated as Illum-III (Experiment III). Examples of images obtained under these conditions are shown in Figure 11. Observing the three figures, some visual differences can be noted due to the different lighting sources. For example, in Figure 10(a), due to the specific lighting system, the illumination is uniform in the entire image. But, in Figures 10(b) and 10(c), a non-uniform illumination can be easily visualized due the room lighting conditions in which the images were acquired.

(a)

(b)

(c)

Figure 11: Some sample images under illumination conditions (a) IL-I (b) IL-II (c) IL-III

Other images with head tilted (Rotation) (Experiment IV) and face occlusion (Experiment V) were acquired under the same three illumination conditions mentioned previously, as shown in Figure 12. Images with scale and noise, mainly blur and color noise, were artificially generated by an image editor with two different levels for each category. For the scale, the size of the images were reduced to 95% (Scale-I, Experiment-VI) and enlarged to 105% (Scale-II, Experiment -VII). Likewise, two different noise levels (I and II) were applied to the images (Blur-I, Blur-II, Color Noise-I, Color Noise-II). These images, in the same order, were used in the experiments VIII to XI. The type of experiments that are classified according to the categories of still images are summarized on Table 1. Some sample images of Blur-II and Color Noise-II are shown in Figure 13.

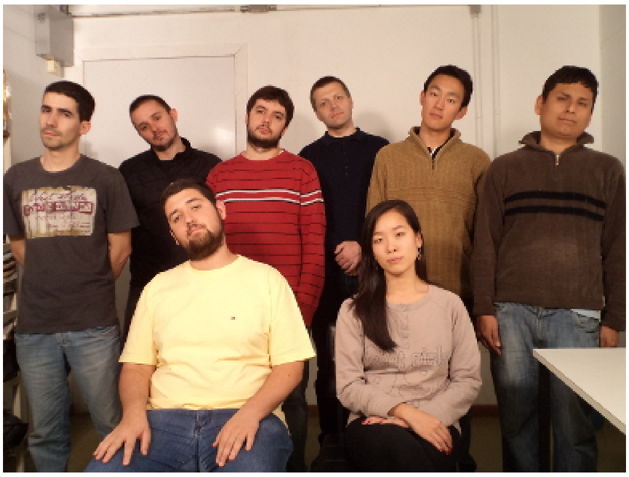
Table 1: Types of experiments according to the categories of still images

Experiment Number	Still Image Category	Still Image Condition
I	ILLUM-I	with two specific lights
II	ILLUM-II	with two specific and room lights
III	ILLUM-III	with room lights
IV	ROTATION	with tilted heads
V	OCCLUSION	with faces partilly occluded
VI	SCALE	with 95% of original size
VII	SCALE	with 105% of original size
VIII	BLUR-1	with blur level I
IX	BLUR-2	with blur level II
X	RGB-1	with noise level I
XI	RGB-2	with noise level II

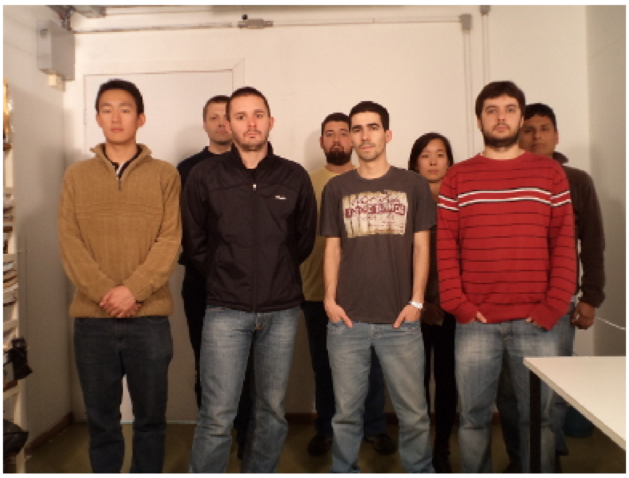
All face object images used in this work were obtained under Illum-I (shown in Figure 14). It is important to emphasize that all face object images (single faces) were obtained separately and are different from those in the images with multiple faces. The size of the images with multiple faces is 2592 x 1944 pixels and the face object images varies from 180 to 270 pixels in width and 240 to 340 pixels in height.

In this work, all the image processing functions were implemented using OpenCV, and the improved ABC algorithm was written using C programming language. All experiments were run on a cluster of computers with Pentium quad-core processors running Linux.

Initially, all parameters of the iABC used in this work were defined empirically, after some preliminary experiments. However, the size of the bee population (SN-80), maximum number of cycles (MCN-100) and the maximum number of runs (NRUNS-30) were also obtained from previous work (CHIDAMBARAM; LOPES, 2010).



(a)



(b)

Figure 12: Some sample images with (a) rotation (b) occlusion

(a)

(b)

(c)

Figure 13: Some sample images with (a)Original Image without noise (b) Part of the image with Blur-II (c) Part of the image with Color Noise-II

Figure 14: Face object images (No. 1 to 20, from top left to right)

4.1.2 PARAMETERS AND THRESHOLDS TUNING

The main goal of the present experiment was tuning of the two parameters of the iABC algorithm and the two thresholds for SURF with which the optimal rate of recognition can be obtained. The iABC parameters are: stagnation (or convergence) factor, and decimation factor (explained in Section 2.3.1). The thresholds refer to T_{CDE} and T_{DDE}, as explained in Section 2.2.1.3. Convergence factor defines when the decimation (or explosion) of a part of population should be done during the optimization process.

Using a pre-defined set of face object images and still images from different conditions, a set of experiments (varying the convergence factor from 10 to 40 and decimation factor 20 to 50, in steps of 10) were conducted to determine the two optimal parameters of the iABC algorithm. For each set of parameters, with still images for each image condition and a face object image, totally twelve experiments were performed. No independent analysis of these parameters was done because the decimation factor depends on the convergence factor. The rates shown in Table 2 represents the average of recognition rate from twelve experiments obtained for each set of convergence and decimation factor. According to the average recognition rates shown in the table, the convergence and decimation factors set (10%, 20%), (10%, 30%) and (10%, 50%) produced the highest values among all other sets. In addition to this, the standard deviation of the set (10 %, 30 %) is lowest among the other selected sets. This is also reflected in Figure 15 (Box Plot Graph) which is obtained using F and p (ANOVA) statistical test. Therefore the set (10%, 30%) was selected to be used as a set of parameters for iABC algorithm in the experiments of MFR.

Table 2: Average recognition rates varying convergence and decimation factors of iABC

Conv.	Decimation Factor (%)							
Factor(%)	20		**30**		40		50	
	Avg.Rec Rate(%)	+Std. Dev.	Avg.Rec Rate(%)	+Std. Dev.	Avg.Rec Rate(%)	+Std. Dev.	Avg.Rec Rate(%)	+Std. Dev.
10	74.74	0.2256	**75.79**	**0.2012**	71.23	0.2585	74.69	0.2355
20	73.68	0.2108	71.75	0.2334	73.68	0.2241	72.28	0.2594
30	71.23	0.2453	72.11	0.2480	72.14	0.2454	69.47	0.2480
40	74.56	0.2223	72.09	0.2427	70.88	0.2296	72.98	0.2489

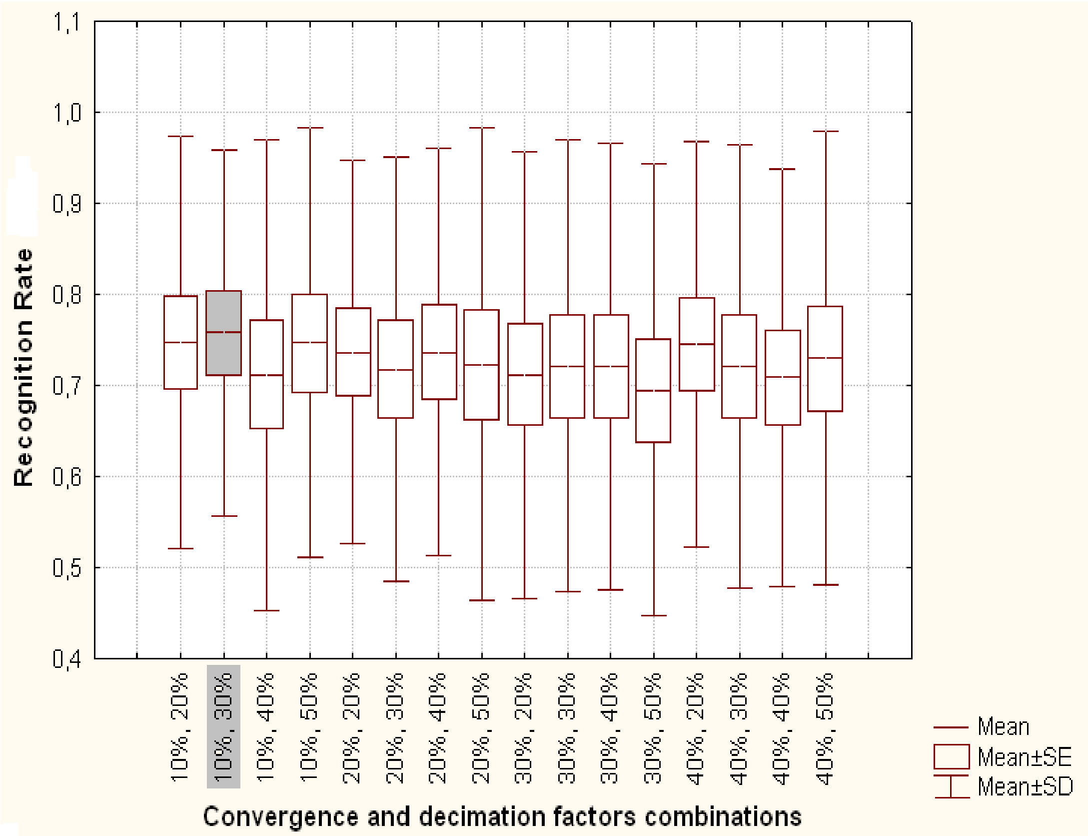


Figure 15: Comparing variations of different combinations of convergence and decimation factors using average recognition rate to determine the optimal thresholds for the iABC algorithm

Similar to iABC parameters tuning, more experiments were done to find the threshold values for coordinate distance and descriptor distance of SURF. Both refer to the Euclidean distance of coordinates and descriptors between two interest points independently. The calculation of these distances is detailed in Section 2.2.1.3. The experiment was done varying the coordinate distance threshold from 20 to 60, in steps of 10 and the descriptor distance threshold from 0.03 to 0.12, in steps of 0.03. Using all sets of thresholds, a total of 240 experiments were performed with still images covering different conditions. The results of the threshold tuning is shown in Table 3 in which the average recognition rates and standard deviation of different combinations of coordinate and descriptor distance are detailed. According to the table and the box plot graph which is obtained using statistical test (shown in Figure 16), it can be noted that the descriptor distance, 0.06, provides some high rates among the other sets of thresholds. Under this descriptor distance, the coordinate distance, 50 provides the maximum rate of recognition and lowest standard deviation. Therefore, the set (50, 0.06) is considered as the suitable threshold for further experiments.

Table 3: Average recognition rates varying coordinate and descriptor distance of SURF

Coordinate	Descriptor distance							
Distance	0.03		**0.06**		0.09		0.12	
	Avg.Rec Rate(%)	±Std. Dev.	Avg.Rec Rate(%)	±Std. Dev.	Avg.Rec Rate(%)	±Std. Dev.	Avg.Rec Rate(%)	±Std. Dev.
20	35.00	0.1845	66.67	0.2412	50.56	0.2039	45.56	0.1966
30	61.11	0.1945	74.44	0.3099	75.00	0.1888	68.89	0.1737
40	69.44	0.2550	85.56	0.1966	80.56	0.1895	67.22	0.2933
50	78.89	0.2324	**90.00**	**0.1348**	77.78	0.2672	69.44	0.2502
60	83.33	0.2137	86.11	0.1895	85.56	0.1903	72.78	0.1874

4.1.3 PRELIMINARY EXPERIMENT WITH SURF-IABC AND MATRIX SURF-IABC

The main objective of this preliminary experiment was to define which approach, SURF-iABC or Matrix SURF-iABC, is the best for the proposed MFR approach. Therefore, both approaches are compared based on the average recognition rate and execution time. This experiment was done using the parameters and thresholds defined in the previous Section (4.1.2) and a set of 20 images from all image conditions. Results are shown in Figure 17, in whick the two middle lines (Avg MatrixSURF-iABC and Avg SURF-iABC) belong to the average recognition rate of all images of both approaches.

As shown in Figure 17, the performance of Matrix approach is higher than SURF-iABC in the major part of the image conditions. The average recognition rate and the average execution time per experiment (a total of 20 experiments) spent during the iterative process of the iABC algorithm are shown in Table 4. The gain regarding the average recognition rate of Matrix iABC-SURF against the traditional approach iABC-SURF is about 9.67%, But, the

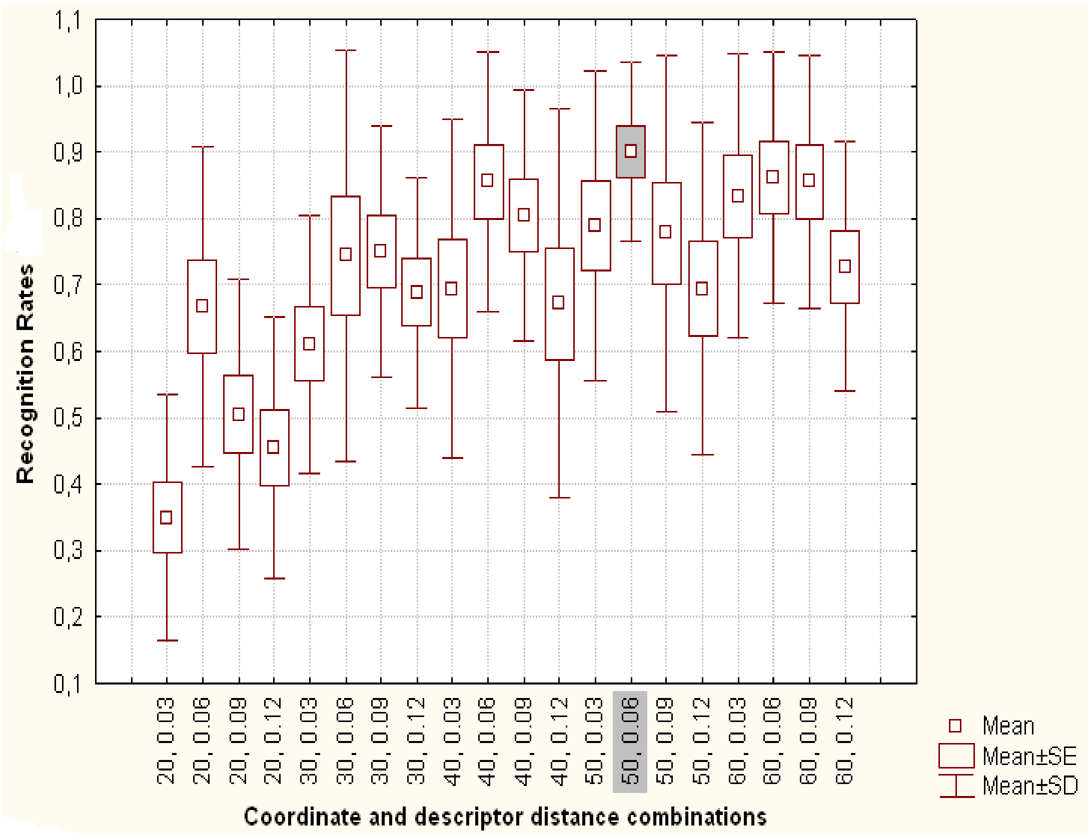


Figure 16: Comparing variations of different combinations of coordinate and descriptor distance using average recognition rate to determine the thresholds of SURF

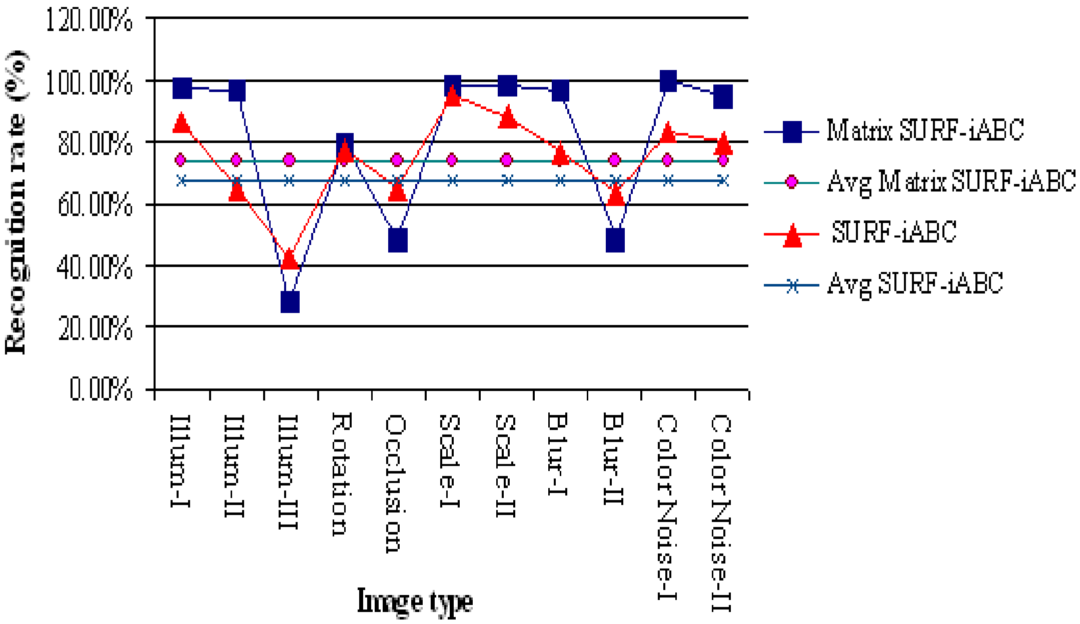


Figure 17: Comparison of recognition rates between SURF-iABC and Matrix SURF-iABC

Matrix approach needs more time than the SURF-iABC. Though the SURF-iABC approach performs well in all image conditions, the Matrix SURF-iABC achieves the highest rates in the major part of the image conditions. Also, a still image with multiple faces captured from real-world conditions may contain one or more types image variations. Moreover, depending on the situation in which the MFR is done, the recognition rate will be relatively more important than a execution time. Based on this context and on the tendency of experimental results, the Matrix SURF-iABC approach can be considered generally better than the traditional SURF-iABC approach. Therefore, the Matrix SURF-iABC was used in the remaining experiments.

Table 4: Average recognition rate and execution time of SURF-iABC and Matrix SURF-iABC approaches

Average Value	SURF-iABC	Matrix SURF-iABC	Gain (%)
Recognition Rate (%)	70.28	77.81	9.67
Execution Time (secs)	494	531	-6.91

Even though the performance of the Matrix SURF-iABC is better than the SURF-iABC approach, the image conditions such as Illum-III, occlusion and blur-II (shown in Figure 17) call attention due to their recognition rates. In these conditions, the Matrix SURF-iABC has no gain against the SURF-iABC approach, as shown in Table 5. This result signals that under conditions like Illum-III, SURF-iABC approach may be appropriate since the images are cut using scale and angle parameters from the still image. It is important to emphasize that Illum-III can appear in real-world conditions from which images are commonly acquired.

Table 5: Comparison of average recognition rate of of SURF-iABC and Matrix SURF-iABC approaches under Illum-III and Blur-II condition

Image Condition	Recognition Rate (%) SURF-iABC	Recognition Rate (%) Matrix SURF-iABC	Gain (%)
ILLUM-III	42.50	28.75	-32.35
BLUR-II	63.33	48.33	-23.69

4.1.4 EXPERIMENT WITH IMAGES UNDER DIFFERENT CONDITIONS

The main objective of the present experiment was to study the robustness of the proposed approach and check whether it can effectively recognize the faces in images obtained under different conditions. The first three experiments (I, II and III) with Illum-I, Illum-II and Illum-III were done using three still images for each condition and twenty face object images. All other experiments (IV to XI) were done with ten face object images. A total of 140 experiments were conducted using 36 different still images as mentioned in Section 4.1.1. The experiments were grouped according to the eleven image conditions. Results are shown in Table 6 which are average rates of recognition and execution time.

Table 6: Average recognition rates and execution time of Matrix SURF-iABC using images under different conditions

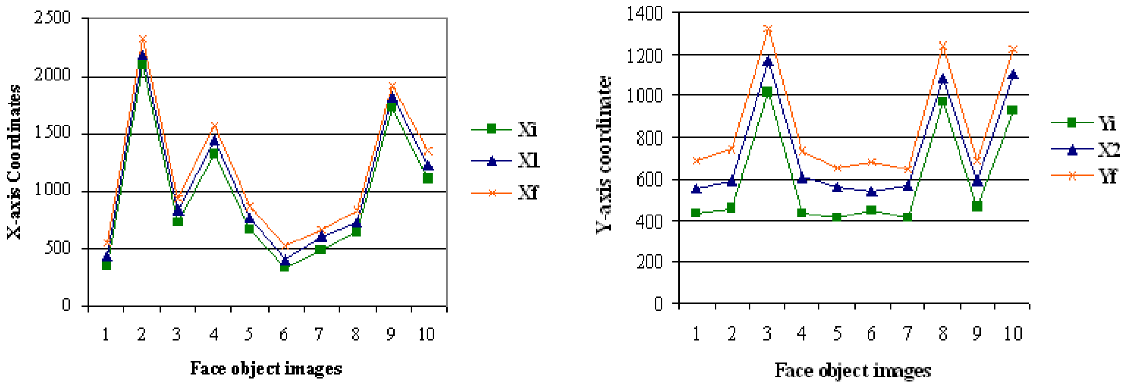
Exp.No.	Experiment Type	Avg. Recognition Rate (%)	Avg. Execution Time per cycle (Secs)
I	Illum-I	81.17	4.14
II	Illum-II	67.33	3.91
III	Illum-III	31.23	3.42
IV	Rotation	51.92	3.69
V	Occlusion	85.33	2.10
VI	Scale-95%	85.67	3.31
VII	Scale-105%	89.00	3.75
VIII	Blur-I	68.67	3.69
IX	Blur-II	47.67	3.82
X	Color Noise-I	85.33	3.78
XI	Color Noise-II	85.00	3.89

Variation generated by different lighting conditions is one of the prominent bottlenecks of face processing tasks. Therefore, several experiments were conducted using still images captured under different lighting conditions. From the results shown in Table 6, a strong influence of lighting conditions on the recognition rate can be observed. The recognition rate decreases gradually from Illum-I to Illum-II and drastically to Illum-III.

In real-world images, it is possible to have occluded faces, mainly, in images with multiple faces. In order to measure the influence of image orientations and to study its impact on the recognition performance, experiments IV to VII were done. But, the images used in experiments were artificially manipulated to vary the scale. As shown in Table 6, the recognition of scale and occlusion is above 85% except for the rotation condition.

The last part of the computational experiments (VIII to XI) is related to the presence of noise and blur in still images. Even with the optimal lighting conditions, the noise and blur that arise from the camera sensor may appear in images (VAN de WEIJER; SCHMID, 2006). Hence, studying the effects of blur and color noise in a face recognition is essential to develop robust approaches. To study the impact of these issues, two sets of experiments was performed using artificially blurred and noisy (color) images. From the experiments, a low recognition rate, below 50%, was obtained for blurred images, more specifically with the Blur-II condition.

In this experiment, in addition to the fitness value which is used to determine the optimal solutions (face images), an additional verification procedure was also implemented to check whether the corresponding central coordinates (X_1, X_2) generated by the iABC algorithm are within the region of the face identified (initial (X_i and Y_i) and final coordinates (X_f and Y_f) in a still image with multiple faces. If they are inside the region, then the recognized face was certainly considered as a valid solution. As an example, the coordinates verification of experiment VII , Scale-105%,is demonstrated, in the plotted graphs of Figure 18(a) for x-coordinate and 18(b) for y-coordinate. The central coordinates of the recognized face is represented by the middle lines in both figures. The initial (X_i and Y_i) and final coordinates (X_f and Y_f) are represented by the other two, top and bottom, lines. In some points, there is no clear separation of lines, i.e, coordinate values are close to each other. This may happen due to the fact that the size of face object image is kept as the same in all experiments independent of the size face present in still images. According to the figures, the central coordinates represented by the middle lines are generally located between the two other lines, which indicates that the present approach can be invariant with scale.



(a) (b)

Figure 18: Coordinates verification of identified face in still images: Central coordinate values (X_1 and X_2) represented by the middle line (a) X-axis initial and final coordinate values (b) Y-axis initial and final coordinate values)

4.1.5 EXPERIMENT WITH FACE OBJECT IMAGES

In Section 4.1.4, the evaluation of the proposed algorithmic approach for still images focused on image conditions (I to XI, as shown in Table 6) was discussed. In the present section, the main goal was to evaluate the robustness of the proposed approach focusing on 20 face object images (shown in Figure 14). Each face was searched in 12 different still images that were not used in the previous experiments. The still images used in this experiment for each face object image covers all image conditions. Thus, a total of 240 (12×20) experiments were done searching for the face object images. Therefore, the result data are grouped according to the face object images. The average recognition rates of face object images are shown in Table 7. The frequency distribution of the recognition rates among all images can be classified as follows: 25% of face object images are between 41 and 60%, 60% of face object images are between 61% and 80% and 15% of face object images are between 81% and 100%. Hence, these results effectively demonstrate the capability of the proposed approach.

Table 7: Average recognition rate of face object images searched in still images under different conditions

Face Obj. Image No.	Avg. Rec Rate (%)
1	67.22
2	70.56
3	56.39
4	67.78
5	86.94
6	88.06
7	63.33
8	66.94
9	69.17
10	71.11
11	46.67
12	75.28
13	60.83
14	62.22
15	84.72
16	63.61
17	42.78
18	41.39
19	67.50
20	41.11

4.1.6 COMPARISON OF RECOGNITION RATES OF ILLUM-I, II AND III

This section discusses the recognition rates obtained with images under conditions Illum-I, II and III from Experiments discussed in Section 4.1.4 and Section 4.1.5. The average recognition rates of all face object images are shown in Table 8, in which the performance of the Matrix SURF-iABC approach decreases from Illum-I to II and II to III. Under this condition, the recognition performance of the approach is low and non-uniform in comparison with other image conditions in all experiments that were conducted so far. It can be noted that the non-uniformity of recognition rates of face images and in some cases, it is close to zero as illustrated in Figure 19.

Table 8: Average recognition rate of face object images in images under illumination I, II and III using Matrix SURF-iABC approach

Face Obj. Image No.	Rec.Rate Illum-I	Rec.Rate Illum-II	Rec.Rate Illum-III
1	62.50	50.00	27.50
2	95.00	48.33	39.17
3	77.50	61.67	47.50
4	85.00	73.33	43.33
5	95.83	79.17	32.50
6	86.67	83.33	41.67
7	95.00	73.33	0.83
8	82.50	73.33	70.00
9	77.50	80.00	55.83
10	93.33	89.17	71.67
11	98.33	50.00	42.50
12	69.17	36.67	22.50
13	40.00	75.83	8.33
14	90.83	90.00	22.50
15	92.50	86.67	81.67
16	95.83	88.33	80.00
17	98.33	60.83	30.83
18	92.50	74.17	35.83
19	49.17	49.17	25.00
20	78.33	93.33	29.17

The average recognition rates of the three illumination conditions I, II and III of both experiments is shown in Table 9, in which, it can observed that the abrupt change of recognition rate under Illum-III. According to the recognition rates of twenty face object images shown in Table 8 and the average recognition rates of the same images regarding the three illumination conditions, it can be observed globally that there is a gradual decrease of recognition rate from Illum-I to Illum-III. But, for some face object images, for example, 9, 13 and 20, the rates does not follow the previous observation. This is more generally due to the uniform illumination of face images even though the image belongs to the Illum-II (no.9 and 20) or Illum-III (no.9).

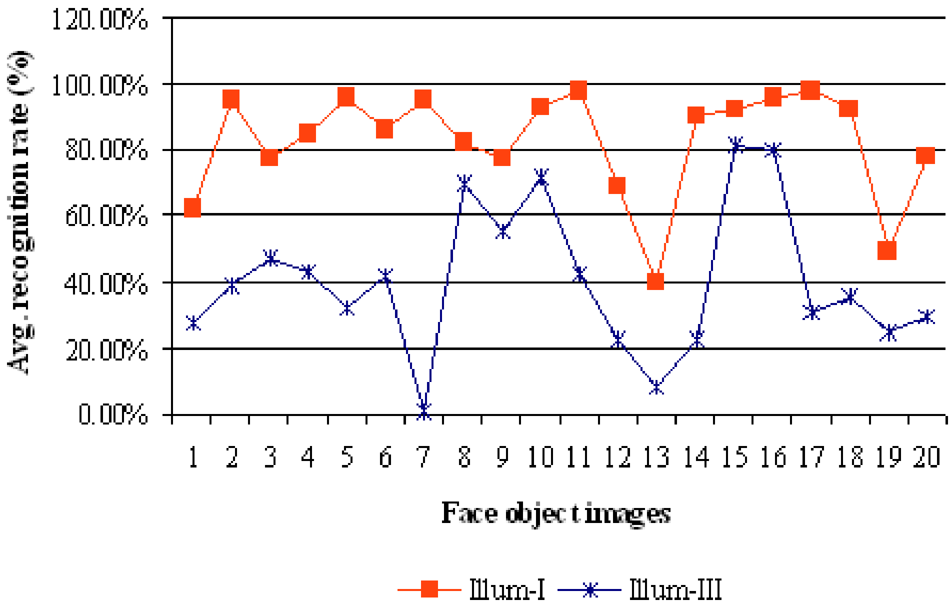


Figure 19: Average recognition rates of Illum- I and III using face object images (Experiment from Section 4.1.5)

Table 9: Comparison of average recognition rates in illumination conditions I, II and III using Matrix SURF-iABC approach

Experiment Type	Illum-I	Illum-II	Illum-III
Image Conditions (I, II, III) from Table 6	80.17	67.33	31.23
Face Object Images (1-20) from Table 8	82.79	70.83	40.42

4.1.7 DISCUSSION OF RESULTS

Within the scope of MFR, to evaluate the robustness of the proposed approach, two major experiments were performed using the still images captured under different conditions. Though the image conditions in real-world environments may be more complex, the still images used in the experiments cover the general image conditions. Since the changes in illumination appear at the top of the list of problems that affect recognition performance (NSTC, 2006) and it still remains a problem for state-of-art algorithms (BEVERIDGE; S.BOLME; A.DRAPER; GIVENS; LUI, 2010), several experiments were performed using still images captured under three different lighting conditions.

The experiment results demonstrate that the Martix SURF-iABC approach can be applicable for the major part of the image conditions. In the first experiment using images under different conditions (Section 4.1.4), the recognition rates degrade under Illum-II, Illum-III, Rotation, Blur-I and II conditions (shown in Table 6). Similarly, in the second experiment using face object images, the same tendency can be observed under Illum-II and Illum-III (shown in Table 8). The brief discussion of Section 4.1.6 emphasizes the influence of illumination variation in the recognition performance.

More specifically, under Illum-III, the high rates generally refer to the faces with uniform illumination, meanwhile low rates are obtained from face images with uncontrolled and partial illumination. This may happen due to the fact that the images from Illum-I and II were obtained using a special lighting system (with high intensity) meanwhile the images from Illum-III were acquired using just room lighting condition (with low intensity). In other words, the Illum-III condition generates faces with non-uniform illumination and shadows according to the position of lights and their reflection on different parts of the image. Hence, the main drawback of the present approach is related to the images with non-uniform illumination condition which leads to uncontrolled variation of intensity of pixels in different regions of still images. Besides, in this kind of images, as suggested by the results, an overall deduction can be done that the SURF can not provide feature descriptors with sufficient discriminative power. From the experiments conducted so far, the illumination variation was identified as the main issue that should be studied deeply. Furthermore, in addition to the SURF method, a reinforcement using other methods may be necessary for enhancing feature extraction purposes. Another possible way to overcome the illumination variation problem could be attempting to use illumination compensation approaches or different combination of feature extraction methods.

The last part of the computational experiments (VIII to XI) was related to the presence of noise and blur in still images. The main influence of blur is on the transition of edges, not on the color change (VAN de WEIJER; SCHMID, 2006). The proposed approach achieves recognition rates above 85% with color noise. On the other hand, the performance degrades with blur.

Besides the illumination conditions, the images under different orientations were tested to evaluate the variability of face recognition approaches. Capturing images at different angles may depends upon the environmental conditions and may be done in some special circumstances. Among different image orientations, the presence of rotation (head tilted images) affects the recognition performance.

From the above analysis, the problems encountered using Matrix SURF-iABC are related to illumination variation, rotation, and blur. In the Matrix SURF-iABC approach, the interest points are calculated only once on the entire still image with multiple faces. Consequently, this approach is unable to consider some image variations, for example, rotation. An alternative solution to overcome the drawbacks of the Martix SURF-iABC may be found using the traditional SURF-iABC associated with other methods such as feature extraction methods based on edges, illumination compensation and preprocessing methods.

4.2 SINGLE FACE RECOGNITION

In the following section, the single face image database and the parameter setting procedures are described first. The experiments of this section are divided into two main categories: Single-stage independent experiments and Three-stage Hierarchical experiments. In the first, the SFR was performed using CA, EHD and SURF independently. In the hierarchical approach, the experiments were done in three-stages through two sequences of (CA, EHD and SURF), and (EHD, CA and SURF).

4.2.1 IMAGE PREPARATION AND EXPERIMENT DETAILS

Face images were obtained from a database constructed in 2011, consisting of 138 different individuals and 13 variations for each individual, including profile and other pose images. Original images were captured under high resolution of 2592 x 1944 pixels. They are obtained under two illumination conditions: controlled (CIL) and uncontrolled (UCIL). For CIL, a specific controlled lighting system was provided meanwhile for UCIL, just the lighting condition provided by the fluorescent lamps of the room was utilized. In the experiments of this section, a subset of seven classes, namely, frontal viewpoint with neutral face (under CIL, UCIL and lateral-light illumination conditions) and frontal viewpoint with changes on the facial expression and scale (both under CIL and UCIL illumination conditions). In the case of lateral-light illumination condition, the illumination was provided only on the left side of the face. These classes and the corresponding labels are summarized in Table 10. Since the CIL-Frontal images are used as the base images, they were captured following the standardization rules provided by (NIST, 2007).

Table 10: Single face image classes

Label	Illumination	Type
CIL_Frontal	Controlled	Neutral
CIL_Expr	Controlled	Facial expression
CIL_Scale	Controlled	Scale
UCIL_LatFrontal	Uncontrolled lateral	Neutral
UCIL_Frontal	Uncontrolled	Neutral
UCIL_Expr	Uncontrolled	Facial expression
UCIL_Scale	Uncontrolled	Scale

Figure 20 illustrates a sample set of images for 10 subjects in which the columns correspond to those of Table 10. To capture images with slight variation in scale, the camera was fixed at the distance of 1.75 meters. The face expression corresponds to the smile patterns.

Figure 20: Sample single face images

After the image acquisition, normalization of face images is a critical issue in FR systems. Many FR methods require normalized face, for example, holistic features approaches. Facial features that are usually normalized include size, orientation and illumination. In a previous work (PRODOSSIMO; CHIDAMBARAM; LOPES, 2012), by detecting eyes, the database of face images was normalized for size and orientation, except for illumination . Similar to the FR works found in the literature, the single face images were cropped into the the size of 550 x 550 pixels. In all experiments, the faces from the first category, CIL-Frontal, are maintained as the base images. Experiments were run on a desktop computer with Intel Pentium IV 2.8MHz processor and 4GB memory running the Linux operating system.

In order to compare the proposed SFR methodology proposed using our database images with others, another set of single face images was obtained from the FEI [1] face database. Among all categories of images, only three frontal face images were selected: (1) frontal with neutral expression, (2) frontal with smiling face expression and (3) frontal with lateral face illumination. According to the information provided by the FEI database, all images were obtained under the same lighting condition. Hence, they are denominated as CIL_Frontal, CIL_Expr and UCIL_Lat_Frontal following the same criteria used for our database image. Some face images were not used since they appeared in different conditions among three selected categories, for example, the same face image with eyeglass in one category and without eyeglass in another category. Therefore, among two hundred images, only 193 were selected preserving the same type of faces in all three conditions. It is important to mention that FEI database contains face images with beards, mustaches and eyeglasses which are not present in our database captured at UTFPR (Federal University of Technology of Paraná). They were normalized, cropped and resized to 510 x 510 pixels. Some sample images are shown in Figure 21.

Figure 21: Sample single face images of FEI database (Column 1 - frontal with neutral expression (CIL_Frontal), Column 2 - frontal with smiling face expression (CIL_Expr), Column 3 - frontal with lateral face illumination (UCIL_LatFrontal))

[1] Available at:http://fei.edu.br/ cet/facedatabase.html

4.2.2 THRESHOLDS SETTINGS

The following experiments were done in order to determine the optimal thresholds for EHD and SURF so that they can achieve high recognition rates.

As discussed in Section 2.2.1.3, when considering the SURF approach, the similarity between two face images is measured based on the repeatability rate. Since considering an exact match between the interest point's coordinates is not feasible, an acceptable tolerance margin must be determined. In fact, two thresholds need to be defined: one related to the coordinate distance error (T_{CDE} - Equation 1) and another to the descriptor distance error (T_{DDE} - Equation 2).

For the experiments, the following intervals of $20 \leq T_{CDE} \leq 70$ and $0.2 \leq T_{DDE} \leq 0.7$ were established. The recognition rates for each combination are shown in Table 11 (the best values are in bold).

Table 11: Recognition rates for different combination of T_{CDE} and T_{DDE} thresholds using SURF with our face database

	T_{DDE}					
T_{CDE}	0.2	0.3	0.4	0.5	0.6	0.7
10	84.06	86.23	84.78	**86.96**	84.78	84.06
20	85.51	**88.41**	**89.86**	86.23	**87.68**	**84.78**
30	**86.23**	88.41	87.68	85.51	84.78	84.78
40	86.23	83.33	86.96	83.33	82.61	76.81
50	85.51	84.78	86.96	84.06	78.26	75.36
60	84.78	84.78	83.33	83.33	77.54	76.09
70	84.06	84.06	83.33	81.88	76.81	73.91

The best results were obtained for $CDE = 20$ for almost all DDE. Therefore, the experiments with our database of face images, $CDE = 20$ and $DDE = 0.4$, which yielded the highest recognition rate, are used as threshold values. The recognition rates shown in Table 11 represent just the ratio of true positive values and total number of face images used in the experiment. There will be no variation if the experiments are repeated. Therefore, no need for statistical analysis to select the final thresholds. This is valid for the following experiments too.

Though the FEI images are somehow similar to our database of face images, we captured them under different lighting condition. Hence, the previous experiment, threshold analysis was done again to determine the thresholds for CDE and and DDE for FEI images. Through some preliminary experiments, the intervals of $05 \leq T_{CDE} \leq 50$ and $0.2 \leq T_{DDE} \leq 0.6$ were established. The recognition rates are shown in Table 12 in which the set T_{CDE} = 10 and T_{DDE} =

0.04 threshold values were considered as the appropriate for experiments with FEI images. The recognition rate of FEI faces is relatively low in comparison with our face images because, the FEI base contains face images with mustache, beard and eyeglass.

Table 12: Recognition rates for different combination of T_{CDE} and T_{DDE} thresholds using SURF with FEI face database

	T_{DDE}				
T_{CDE}	0.2	0.3	**0.4**	0.5	0.6
05	28.50	31.61	31.09	30.57	31.61
07	38.86	42.49	44.56	48.19	47.67
10	49.22	53.89	**55.96**	53.37	50.78
20	48.19	48.70	49.74	48.19	47.67
30	48.19	46.63	44.04	44.56	39.90
40	47.67	45.60	43.01	43.52	36.79
50	47.15	45.60	41.45	41.45	34.72

In Section 2.2.3, there is an explanation about the construction of EHD. In summary, the occurrence frequency of five different edge types is determined from both the combination of uniform image blocks and the whole image, thus generating three edge histograms: local, semi-global and global. However, an image block contributes to the histogram only if the maximum strength value of any edge type is greater than a threshold value, denoted by T_{edge}.

The value that yields the best recognition rates was determined based on a coarse to fine tuning considering descriptors composed by features extracted from different types of image blocks. First, the recognition rates for an edge strength in the range $10 \leq T_{edge} \leq 40$ was analyzed. The recognition rates are shown in Table 13. The G, SG and L mean global, semi-global and local features, respectively.

Table 13: Recognition rates for varying T_{edge} using EHD (coarse tuning) with our face database

Features	10	20	30	40
G_SG_L	34,78	31,88	18,84	13,04
G	3.62	2.90	2.90	1.45
SG	32.61	28.99	23.19	20.29
L	**48.55**	41.30	28.99	28.26
G_SG	21.01	13.77	10.87	7.97
G_L	28.26	23.19	13.77	12.32
SG_L	45.65	40.58	30.43	27.54

In the EHD threshold analysis, the highest rate was obtained using local features and $T_{edge} = 10$. Since the values decrease as the threshold increases, a fine tuning range was defined as $3 \leq T_{edge} \leq 15$. The results are shown in Table 14.

Table 14: Recognition rates for varying T_{edge} using EHD (fine tuning) with our face database

Features	3	5	8	10	12	15
G_SG_L	25.36	.28.26	30.43	34.78	28.26	32.61
G	3.62	5.80	3.62	3.62	2.90	3.62
SG	27.54	36.23	33.33	32.61	32.61	28.99
L	**51.45**	49.28	46.38	48.55	44.93	39.13
G_SG	15.22	17.39	18.12	21.01	15.94	17.39
G_L	21.74	23.19	26.09	28.26	22.46	28.99
SG_L	47.10	47.10	45.65	45.65	42.75	39.86

It can be observed from both threshold experiments that the results obtained with local features overcome those obtained with the other combinations of features (G, SG, etc.) for all thresholds. From the fine tuning analysis, the overall best recognition rate was obtained using the edge strength $T_{edge} = 3$. Hence, the experiments discussed in the next section were conducted using the parameters $CDE = 20$, $DDE = 0.4$ (SURF) and $T_{edge} = 3$ (EHD) with our database images.

Similar to the previous experiment in which the threshold for EHD was determined using our database images, the same was also established for FEI images, as shown in Table 15. Regarding FEI face database, no significant results were obtained from coarse tuning. Hence, only fine tuning was done. according to the recognition rates obtained using different values, the threshold 02 was selected.

Table 15: Recognition rates for varying T_{edge} using EHD (fine tuning) with FEI face database

Features	1	2	3	4	5
G_SG_L	3.11	20.21	30.05	11.40	7.77
G	1.55	3.11	7.25	3.63	0.00
SG	2.07	21.24	27.46	11.92	8.29
L	4.66	**47.15**	41.97	16.06	10.36
G_SG	2.59	6.22	18.13	7.25	5.18
G_L	3.11	15.54	27.46	10.36	6.22
SG_L	3.63	39.90	38.86	15.03	9.33

All threshold values obtained for SURF and EHD experiments with our face database and FEI face database are summarized in Table 16.

Table 16: Threshold values for SURF and EHD

Features Combination	SURF T_{CDE}	SURF T_{DDE}	EHD T_{edge}
Our face database	20	0.04	3
FEI face database	10	0.04	2

4.2.3 SINGLE-STAGE INDEPENDENT EXPERIMENTS

The main objective of the single-stage experiments is to evaluate the CA, EHD and SURF methods individually. Therefore, it is possible to determine the discriminative power of features represented by color angles, edge histograms, and interest point descriptors when the query image is acquired under different image conditions (related to illumination, facial expressions and scale). It is important to recall that the search database contains only images that belong to the CIL_Frontal class. The remaining samples, here denoted by CIL_Frontal, CIL_Expr, CIL_Scale, UCIL_LatFrontal, UCIL_Frontal, UCIL_Expr and UCIL_Scale (Table 10), are used as query images. Besides, an additional experiment considering all 828 images (six classes of 138 samples) as queries was also performed (denoted by ALL-Images in the following sections) to study the performance of the three methods.

The feature extraction using CA follows the same methodology of EHD. The application of CA and EHD approaches allows to assess the impact of considering local, global and semi-global features. Tables 17 and 18 show the recognition rates for each scenario (the best values are marked in bold).

Table 17: Recognition rates (%) using CA with our face database

Features Comb.	CIL_ Expr	CIL_ Scale	Mean CIL	UCIL_ LatFrontal	UCIL_ Frontal	UCIL_ Expr	UCIL_ Scale	Mean UCIL	All_ Images
G+SG+L	65.22	**76.09**	**70.65**	**10,14**	21.01	**15.22**	17.39	**15.94**	**34.18**
G	25.36	23.91	24.64	0,72	3.62	2.90	2.17	2.36	9.66
SG	65.22	73.91	69.57	2.90	15.22	8.70	13.04	9.96	29.95
L	42.03	65.94	53.99	6.52	23.19	13.77	15.94	14.86	27.78
G+SG	**66.67**	73.19	69.93	4.35	11.59	9.42	13.77	9.78	29.59
G+L	46.38	70.29	58.33	6.52	**25.36**	**15.22**	**21.01**	17.03	30.68
SG+L	61.59	75.36	68.48	**10.14**	24.64	**15.22**	18.84	17.21	34.06

Regarding the CA approach, it can be noted that there is no specific combination of features that yields the highest recognition rate for all image classes (Table 17). For images taken under CIL, the combinations considering semi-global features lead to the best results (see

Table 18: Recognition rates (%) using EHD with our face database

Features Comb.	CIL_ Expre	CIL_ Scale	UCIL_ LatFrontal	UCIL_ Frontal	UCIL_ Expre	UCIL_ Scale	All_ Images
G+SG+L	25.36	8.70	8.70	28.26	14.49	5.07	15.10
G	3.62	2.17	1.45	2.90	3.62	1.45	2.54
SG	27.54	13.04	10.87	34.78	14.49	13.04	18.96
L	**51.45**	**20.29**	**28.99**	**51.45**	**26.81**	**28.99**	**34.66**
G+SG	15.22	2.90	2.90	12.32	5.80	1.45	6.76
G+L	21.74	6.52	5.80	21.74	10.14	2.90	11.47
SG+L	47.10	18.84	20.29	48.55	22.46	22.46	29.95

Table 17 - column Mean CIL). On the other hand, for UCIL, the results indicate that local features were the most efficient (see Table 17 - column Mean UCIL). Also, the global features lead to poor results for all sources of variations, thus confirming their limited discriminative power. Based on such observations, the combinations SG+L and G+SG+L can be chosen as the best options.

When considering the EHD approach, the highest recognition rates are always obtained for descriptors composed simply by local features (L) (Table 18). In each category of face images, the local features always highlights among other combinations.

Unlike CA and EHD, SURF is defined based only on local features. Hence, there is no need for feature combinations. The recognition rates are shown in Table 19.

Table 19: Recognition rates (%) using SURF with our face database

Features Comb.	CIL_ Expre	CIL_ Scale	UCIL_ LatFrontal	UCIL_ Frontal	UCIL_ Expre	UCIL_ Scale	All_ Images
L	89.13	95.65	69.57	83.33	32.61	62.32	72.10

Figure 22 summarizes the best results obtained from different features combination. For EHD, the combination L is selected as the best. Since the SG+L features alone provide a recognition rate very close to G+SG+L combination and G features are not sufficient enough to achieve a good rate, SG+L feature combination was selected as the best for CA.

The matching process considering SURF descriptors leads to the highest recognition rates in all cases. The CA approach yielded similar results for the UCIL_Frontal, UCIL_Expr and UCIL_Scale classes, thus showing robustness for changes in scale and expression. On the other hand, this method has shown to be sensitive to the most significative illumination variations, as is the case for UCIL_LatFrontal. More generally, the CA and SURF have similar behaviors. Both CA and SURF present lower recognition rates for images taken under UCIL

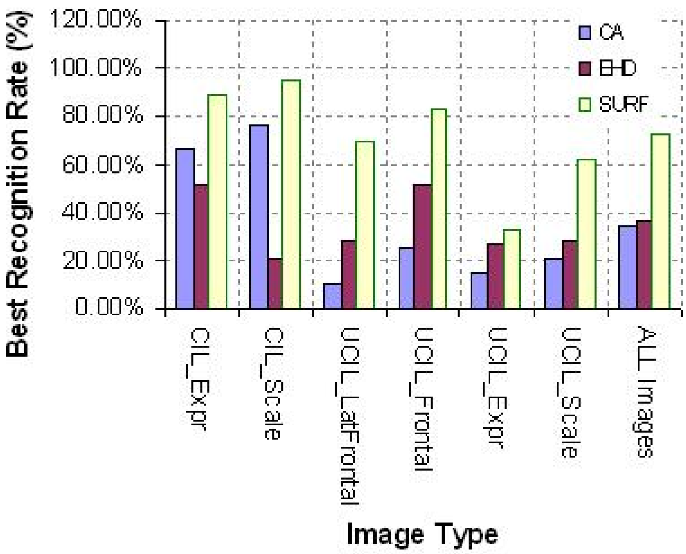


Figure 22: Recognition rates using CA (SG+L), EHD (L) and SURF independently with our face database

conditions. The worst rates when considering SURF were obtained for the UCIL_Expr class, which is justified by the fact that the face image undergoes with two strong variations, illumination and expression.

The results obtained with the EHD show that this approach is less sensitive to changes in the illumination conditions. This can be observed from the recognition rates of the classes CIL_Expr and UCIL_Frontal. The same happens to the CIL_Scale, UCIL_LatFrontal, UCIL_Expr and UCIL_Scale variations.

The execution time shown in Figure 23 takes into account the processing time of all tasks, including image matching, feature extraction and file operations. The times were measured on a desktop with a 2.8MHz Intel Pentium IV processor and 4GB memory. The SURF method takes about twice the execution time of CA. Among the three methods, EHD demands the lowest computational time. In the case of SURF, a slight variation of time is detected in conditions such as CIL_Expr, UCIL_Expr and UCIL_Frontal due to the increase of number of interest points.

With the objective of comparing our proposal with other face images, the single stage independent experiments were conducted using FEI face database. As shown in Table 20, the CA obtains the high rate (highlighted values) with the presence of SG features. On the other hand, Table 21 shows the average recognition rates obtained with EHD for three categories of images in which local (L) features are discriminative than others. With FEI face images, using CA and EHD, the recognition rates are almost similar. The results obtained from SURF are summarized in Table 22. The overall execution time shown in Figure 24 demonstrates the low

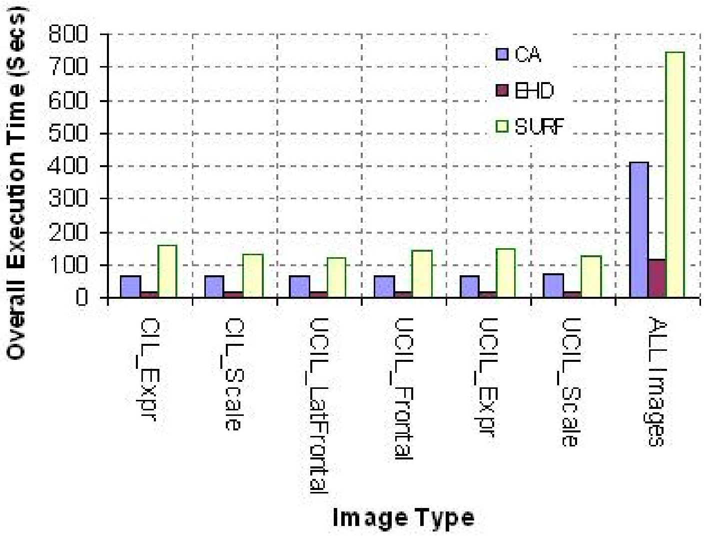


Figure 23: Overall execution time using CA, EHD and SURF independently with our face database

time consumed by EHD.

Table 20: Recognition rates (%) of using CA with FEI face database

Features Comb.	CIL_ Expre	UCIL_ LatFrontal	All_ Images
G+SG+L	46.63	**47.15**	**49.71**
G	18.13	5.70	14.77
SG	**52.33**	27.46	46.63
L	25.39	31.61	29.02
G+SG	49.74	25.39	44.04
G+L	32.64	35.75	35.75
SG+L	44.04	45.08	47.15

Table 21: Recognition rates (%) of using EHD with FEI face database

Features Comb.	CIL_ Expre	UCIL_ LatFrontal	All_ Images
G+SG+L	20.21	29.53	16.84
G	3.11	7.77	3.11
SG	21.24	24.35	16.58
L	**47.15**	**39.38**	**43.52**
G+SG	6.22	17.62	6.74
G+L	15.54	22.80	12.95
SG+L	39.90	35.75	29.27

Table 22: Recognition rates (%) of using SURF with FEI face database

Features Comb.	CIL_ Expre	UCIL_ LatFrontal	All_ Images
L	55.44	49.34	52.39

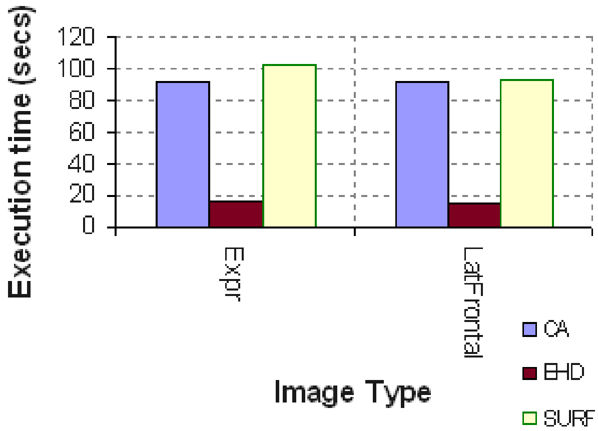


Figure 24: Overall execution time using CA, EHD and SURF independently with FEI face database

4.2.4 THREE-STAGE HIERARCHICAL EXPERIMENTS

In this section, a hierarchical approach was evaluated such that the feature extraction and matching procedures were performed in three stages. More details about this approach was discussed in Section 3.2 and its schematic diagram is shown in Figure 10. One of the main motivations to propose this kind of approach is to achieve a high recognition rate with low computational effort. This is done in two ways. First, the methods that require low computational cost were implemented at the initial stages. Second, only the query images with incorrect matches in a previous stage were processed in the next. The incorrect matches were identified using the name of face image to check whether really the face image classified was correct. To evaluate the method applied in each stage, both execution time and recognition rates were analyzed. In the following experiments, the same image database of the previous section was used.

Since the present proposal is focused on hierarchical approach, only two combinations of three stage experiment were performed. Two experiments were done by keeping SURF always in the third stage, due to its high computational cost, and changing CA and EHD in the first and second stages, alternatively. As stated in the single-stage experiment, the execution time for SURF is the highest among all methods.

Based on the results presented in the previous section, EHD and CA approaches considered descriptors extracted from local (L) and semi-global plus local (SG+L) image blocks, respectively. Table 23 summarizes the accumulated recognition rates of the sequences CA-EHD-SURF and EHD-CA-SURF.

Table 23: Accumulated Recognition rates (%) of the hierarchical approach with our face database

Image Class	CA-EHD-SURF			EHD-CA-SURF		
	CA	EHD	SURF	EHD	CA	SURF
CIL_Expr	61.59	72.46	91.30	51.45	73.91	89.86
CIL_Scale	75.36	79.71	97.10	20.29	79.71	97.10
CIL_LatFrontal	10.14	36.23	81.16	28.99	34.78	78.99
UCIL_Frontal	24.64	61.59	90.58	51.45	62.32	92.75
UCIL_Expr	15.22	35.51	61.59	26.81	35.51	56.52
UCIL_Scale	18.84	39.86	76.81	28.99	40.58	77.54

As shown in Table 23, CA descriptors are more discriminative than EHD when they are considered in the first stage for images acquired under CIL. For the other cases, EHD presents the best recognition rates, confirming the results obtained when analyzing each method separately.

The overall recognition rates of the CA-EHD-SURF sequence were slightly better. However, its average execution time was 20% higher, which can be explained by the higher computational cost of the CA approach, discussed in the previous section (Figure 23). Both sequences present better results than SURF for all sources of variation being considered, as illustrated in Figure 25.

The average recognition rates of all image types of three approaches, SURF, CA-EHD-SURF and EHD-CA-SURF, calculated from the values of Tables 19 and 23, is shown in Figure 26. The average recognition performance confirms the superiority of the hierarchical approach.

It can be observed that the impact of considering the hierarchical approaches is more representative for images taken under UCIL than under CIL conditions. The results of CA and EHD of UCIL condition in both sequences show that the EHD increments the recognition rates significantly. As shown in Table 23, the recognition rates of face images with expression and scale decreases significantly from CIL to UCIL conditions. The main difference between these images is the lighting condition.

The overall execution time of SURF (without ALL_images) is about 138.83 secs, meanwhile for CA-EHD-SURF and EHD-CA-SURF (without ALL_images) is about 146.67 and 122.67 secs, respectively. At the same time, the execution time of the best feature combination of all image conditions can be seen in Figure 27.

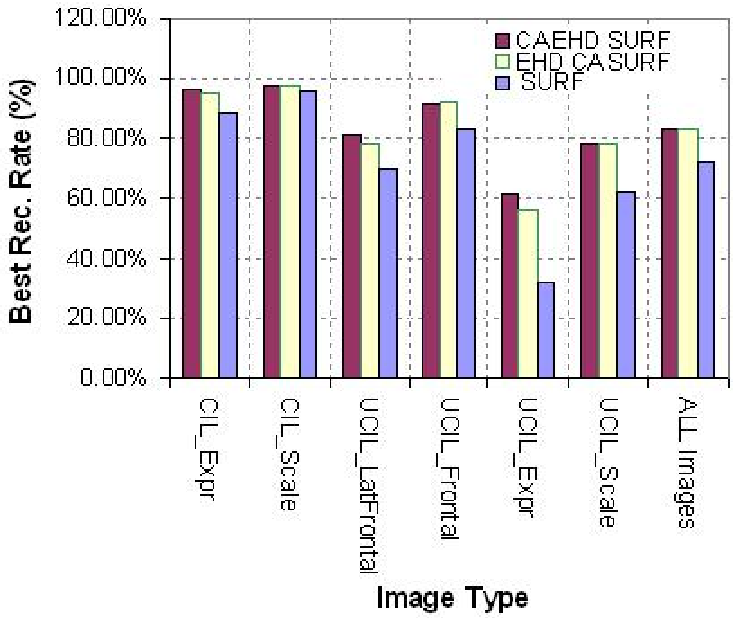


Figure 25: Recognition rate of SURF and three stage hierarchical FR experiments with our face database

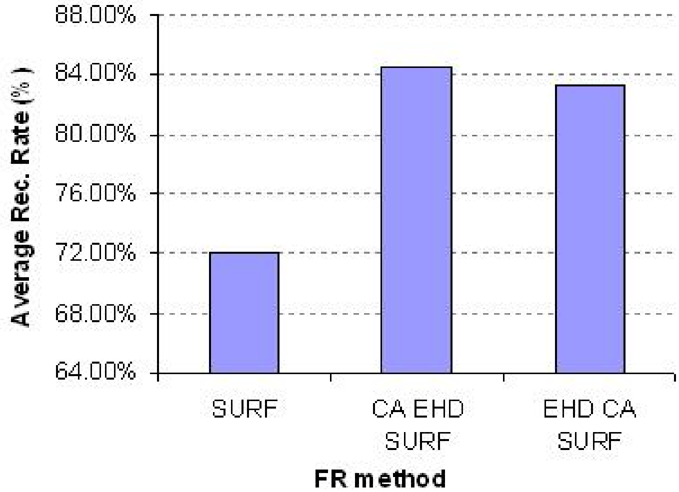


Figure 26: Average recognition rates of SURF and hierarchical FR methods with our face database

Based on these results, the combination of color and edge characteristics with interest points improves significantly the recognition rates. Due to the lowest computational cost, the EHD-CA-SURF sequence can be chosen as the appropriate method for SFR using the our face images captured for this purpose.

The average recognition rates obtained using FEI face images are shown in Figure 28 in which the superiority of the hierarchical three-stage approach regarding single-stage experiment SURF. Between the two hierarchical experiments CA-EHD-SURF and EHD-CA-SURF, the

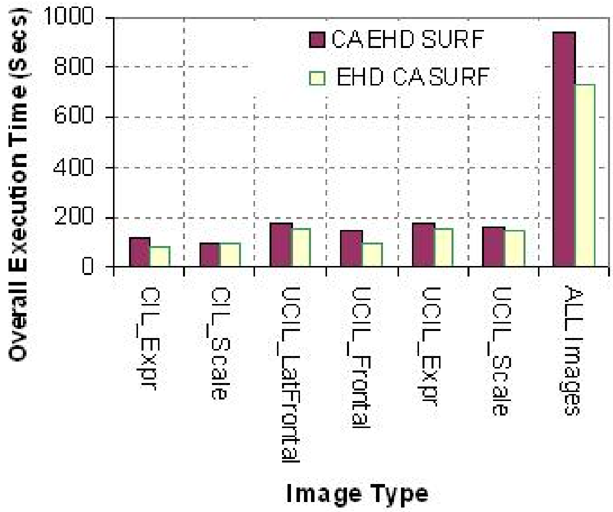


Figure 27: Overall execution time of three stage hierarchical FR experiments with our face database

first sequence beginning with CA come in first place. However, the difference both sequences is small. The overall execution time is shown in Figure 29.

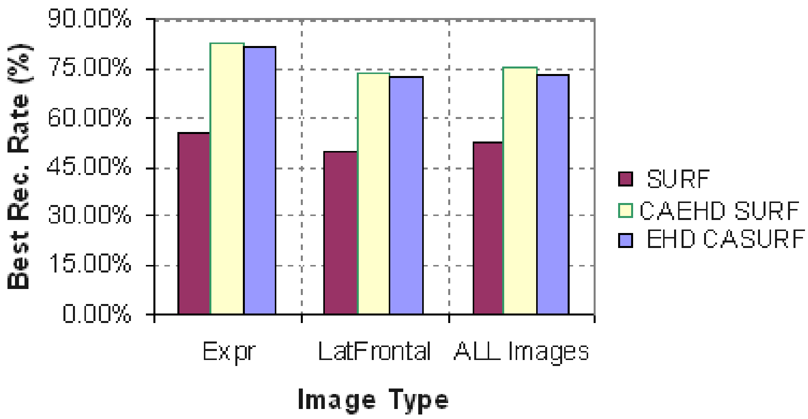


Figure 28: Average recognition rates of SURF and hierarchical FR methods with FEI face database

The recognition rates of hierarchical experiments with our database and FEI database are shown in Figure 30. There is a small difference of 4.88% recognition rate between both face database in the sequence CA-EHD-SURF. However, it can be seen some similar behavior in both case.

4.2.5 DISCUSSION OF RESULTS

The FR approach proposed is based on the global, semi-global and local features of face images. The influence of these features were investigated in the experiments. However,

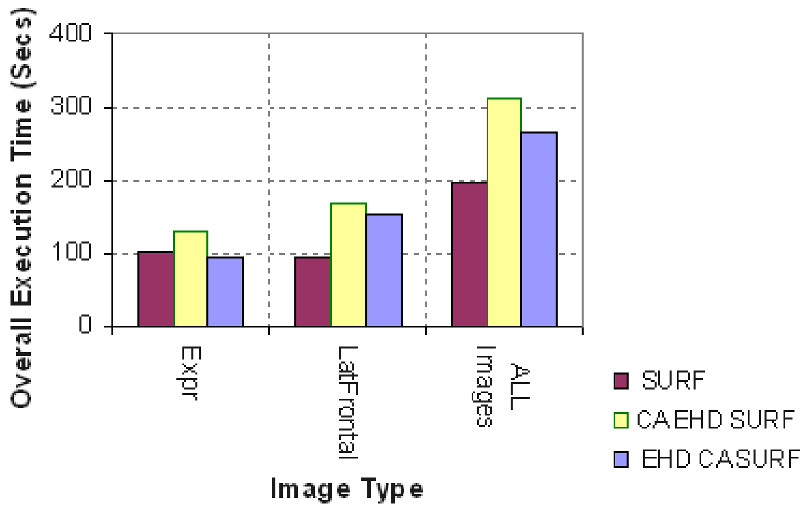


Figure 29: Overall execution time of SURF and hierarchical FR methods with FEI face database

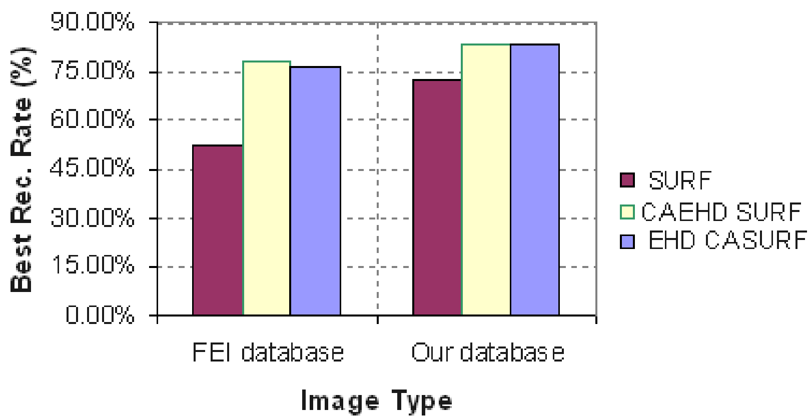


Figure 30: Average recognition rates of three stage hierarchical FR experiments with our face database and FEI face database

the variation of illumination is one of the prominent issues for face recognition tasks. Because the variation of illumination is still considered as the main issue, the single-stage as well as three-stage FR experiments were done on face images obtained under two different lighting conditions. This fact was demonstrated by both approaches under CIL and UCIL images. Among the three methods, CA was not invariant to illumination variation. This may happen due to the fact that, to calculate the color angles, only low order moments (average pixel values of the color channels) were used. One way to overcome issue may be to calculate angles using color edges (FINLAYSON; CHATTERJEE; FUNT, 1996).

From the independent analysis, it can be concluded that SURF is the best for all image conditions, though its high computational cost, when compared to CA and EHD. In the independent analysis of experiments, the performance of CA and EHD highlights the complementary power of features. CA was used because the image features can be represented with a set of few parameters in a compact and representative form. A drawback of this method is that the results are inadequate to demonstrate invariability for illumination variation. Similar to CA, SURF also achieved low rates under UCIL, mainly with face expression. In this condition, the performance was affected since there are two strong changes in images due to expression and illumination change. Consequently, the gradient information may be affected due to the image variations. On the other side, EHD demonstrates to be relatively invariant to illumination variation. The reason for this fact may be that edges tend to be insensitive to a range of illumination conditions. Additional advantage of the EHD is the execution time which is the lowest among SURF and CA. From the three-stage hierarchical experiment, in terms of the recognition rate, the CA, EHD and SURF sequence performs better than the EHD, CA and SURF sequence, though the difference in performance between both sequences was small. Also, in terms of the execution time, the EHD, CA and SURF sequence approach requires low computational effort.

According to the experiments conducted with FEI face database, the results show that the proposed hierarchical approach performs similarly as in our face database. The difference of recognition rate between both databases should be due to the presence facial characteristics such as beard, mustache and eyeglass in the face images of FEI database. These characteristics are not present in our database.

Furthermore, the results obtained from the experiments suggests that the three-stage hierarchical approach performs better than the single-stage using SURF. The analysis of result data show that there is no strong influence of the global features individually in any one of the methods evaluated in the present approach. It is important to emphasize that the SFR approach is strictly based on extracted features from face images. No machine learning algorithms are involved in this process.

5 CONCLUSIONS

In the following sections, the overall conclusions are detailed based on the proposed approach and its experimental results and analysis. The conclusions are divided in two sections: MFR and SFR. Finally, the main contributions and future works are outlined.

5.1 MULTIPLE FACES RECOGNITION

The entire MFR approach was developed based on the fact that, in real-world conditions, the still images with multiple faces are normally captured with a complex background under varying conditions. Hence, the construction of invariant FR approaches to these conditions was become the main objective of the present research. Besides, many studies on images with single faces have been performed for the past two decades. However, very few works deal with multiple faces, and this was one of the main motivations of this work.

The first step was the acquisition of still images with different conditions. Since, the images consist of many variations, the MFR was proposed as an optimization problem. Therefore, the attempts were done with SURF-iABC and Matrix SURF-iABC. Though the SURF-iABC seems to be suitable for MFR at the initial stage, the small difference of performance in the results suggested the Matrix SURF-iABC as the most appropriate method. To evaluate the robustness of the semi-supervised approach with local features, an extensive experimental analysis was done focusing on still images under different conditions and, also, on face object images. In both experiments the same tendency of results was observed. Although, the Matrix SURF-iABC was robust enough for most conditions, the performance of the approach was low, mainly, under the illumination variation condition. The conditions observed in the still images were related to the non-uniform and partial illumination on faces with low intensity of pixels. A variety of images acquired with multiple faces was used to evaluate the robustness of the proposed approach.

As discussed throughout the study, the entire work was performed as a semi-supervised approach using just the discriminative power of local features. The variation of recognition performance influenced by different image conditions was widely discussed in the work. Despite of some drawbacks encountered, an overall conclusion can be done that the semi-supervised SURF-iABC is suitable in MFR under the image conditions involved in this work.

5.2 SINGLE FACES RECOGNITION

The SFR has received significant attention in the past few decades, and at the same time, several techniques have been developed to improve the FR performance. However, with the development of new technologies, FR research challenges still remain. Most of them are related to the efficient processing of images acquired under uncontrolled conditions, where image variation due to illumination conditions, pose or facial expressions are present. Based on this context, a SFR approach was developed and its performance was evaluated under different image conditions.

The semi-supervised FR approaches are not common in the literature and also, they can be conducted with low computational effort in comparison with supervised approaches. Therefore, the present attempt was made to check whether a semi-supervised method can provide high rates of recognition. The results obtained from the experiments confirmed that the semi-supervised approaches can also generate high recognition rate.

In SFR, the development of a semi-supervised hierarchical approach was mainly based on the fact that it can take advantage of different types of features obtained by three image descriptors: CA, EHD and SURF. Thus, it is possible to combine the discriminative power of information related to color, edges and interest points. The entire approach was evaluated considering the features extracted from global, semi-global and local partitioned images.

To evaluate the robustness of each descriptor, first, an independent analysis of FR was performed. The results showed that local features are the most significative to describe characteristics associated to edges, on the other hand, semi-global combined with local features are more representative for color information. No predominant influence of global features was detected from the experimental results. Although SURF presented the best results for all image conditions, its computational cost is higher than the other methods.

A hierarchical three-stage approach was proposed in order to reduce the computational cost, while taking advantage of the complementary information of the different descriptors. In the approach, the feature extraction and matching procedures were performed in three stages. In summary, only the query images that were not matched in previous stages were processed in the following stages. In this context, as suggested by the results, the sequence EHD, CA and SURF was considered as the most appropriate approach for the type of images involved in the experiments. Overall, the hierarchical approach provided higher recognition rates in comparison with SURF, individually. According to the experiments conducted on our face database and FEI face images, it is evident that the hierarchical approach gradually improves the recognition rate each and every stage. Similar to the MFR approach, in SFR also, no training or learning was done. This is one of the main differential of this work regarding the others in the literature.

5.3 CONTRIBUTIONS

The main contribution of the FR work comes from the construction of novel approaches to recognize faces in still images with multiple faces and the recognition of single faces from a database of images under varying conditions through semi-supervised approach. Some other contributions are summarized in this section.

- From the MFR conducted in this work, it is possible that treating the search for interest regions and recognition of faces as an optimization process is one of the directions to cope with image variations.
- The hierarchical architecture developed for SFR can considered as an another main contribution. The hierarchical approach proposed with the combination of global, semi-global and local features was proved to be an good option for SFR. In addition to this, this kind of approach can work with low computational cost and the feature extraction methods can be easily changed or substituted by others. The contribution of this part is that the proposed approach can be applied to any kind of images.
- Regarding features combination, it is also possible that the local features extracted from edges and the semi-global and local features extracted from color information are the most discriminative in SFR.

More generally, it can be said that the proposed methodology for MFR and SFR are not common in the literature. Also, another specific contribution comes from the construction of single data faces with more than 4500 images (JPEG format) which will be available for the international research community.

5.4 FUTURE WORK

In MFR, it has been shown that the main issues are rotation, blur, and illumination variation. The presence of non-uniform illumination and change of lighting condition from one image to another degrades the recognition performance when using the proposed approach. Hence, the variation of illumination appears as a main issue and remains an open subject of research. Though there are many directions for further research, future works will be focused on the illumination variation problem to improve the overall performance of the proposed approach. More than 380 experiments using different still images and face object images were performed. However, for better understanding and to propose more precise methods for MFR, the experimental analysis may be focused on individual conditions. The present investigation showed that SURF is robust enough under many image conditions. But, the FR performance is inadequate under some specific conditions as shown by experiment results. Therefore, new attempts could be done using other feature extraction methods with SURF.

In SFR, in comparison with SURF, the effective performance of EHD as well as CA were not sufficient under most of the image conditions involved. Regarding CA, future works may include attempts to increase the recognition rate by considering the edge information. On the other side, in EHD, the edges are the main source of information. Hence, the performance EHD can be improved by making edges more evident. Furthermore, additional experimental analysis with other base images will be another direction to compare with the results obtained in the present proposal. In this work, only one size of sub-images was studied based on the procedure of EHD in SFR. It is also quite interesting to study the feature extraction methods with other sizes of sub-images.

More generally, considering the discussion of results, it can be noted that the performance of the FR approach can be improved in both MFR and SFR, by applying some pre-processing or illumination compensation techniques. To improve the performance of the approach from the current level, another possible direction could be the extraction of complementary features using some other feature-based methods which are invariant to different lighting conditions.

REFERENCES

ABATE, A. et al. 2D and 3D face recognition: A survey. **Pattern Recognition Letters**, v. 28, p. 1885–1906, 2007.

AHONEN, T.; HADID, A.; PIETIKAINEN, M. Face description with Local Binary Patterns: Application to face recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 28, n. 12, p. 2037–2041, 2006.

AMARAL, V. do; THOMAZ, C. E. Extração e comparação de características locais e globais para o reconhecimento automático de imagens de faces. In: **Anais do VIII Workshop de Visão Computacional**. [S.l.: s.n.], 2012.

ARANDJELOVIC, O.; CIPOLLA, R. A methodology for rapid illumination-invariant face recognition using image processing filters. **Computer Vision and Image Understanding**, v. 113, p. 159–171, 2009.

ASBACH, M.; P.HOSTEN; UNGER, M. An evaluation of local features for face detection and localization. In: **Proceedings of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services**. [S.l.: s.n.], 2008. p. 32–35.

AZAD, P.; ASFOUR, T.; DILLMANN, R. Combining Harris interest points and the SIFT descriptor for fast scale-invariant object recognition. In: **Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems**. [S.l.: s.n.], 2009. p. 4275–4280.

BAHAMISH, H. A. A.; ABDULLAH, R.; SALAM, R. A. Protein conformational search using bees algorithm. In: **Proceedings of the Second Asia International Conference on Modelling & Simulation**. [S.l.: s.n.], 2008. p. 91–916.

BAY, H. et al. Speeded-Up Robust Features (SURF). **Computer Vision and Image Understanding**, v. 110, n. 3, p. 346–359, 2008.

BELHUMEUR, P.; HESPANHA, J.; KRIEGMAN, D. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 7, p. 711–720, July 1997.

BELHUMEUR, P. N.; KRIEGMAN, D. J. What is the set of images of an object under all possible illumination conditions? **International Jornal of Computer Vision**, v. 28, n. 3, p. 1–16, 1998.

BEVERIDGE, J. et al. Quantifying how lighting and focus affect face recognition performance. In: **Proceedings of the Computer Vision and Pattern Recognition Workshop**. [S.l.: s.n.], 2010. p. 74–81.

BONABEAU, E.; DORIGO, M.; G.THERAULAZ. **Swarm Intelligence: From Natural to Artificial Systems**. [S.l.]: Oxford University Press, 1999.

BRUNELLI, R.; POGGIO, T. Face recognition: Features versus templates. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 15, n. 10, p. 1042–1052, 1993.

CAMPADELLI, P.; LANZAROTTI, R.; LIPORI, G. Automatic facial feature extraction of face recognition. In: ____. **Recent Advances in Face Recognition**. Rijeka, Croatia: InTech Education and Publishing KG, 2008. cap. 3, p. 32–58.

CHELLAPPA, R.; WILSON, C.; SIROHEY, S. Human and machine recognition of faces: A survey. **Proceedings of the IEEE**, v. 83, n. 5, p. 705–741, may 1995.

CHIDAMBARAM, C.; LOPES, H. S. An improved artificial bee colony algorithm for the object recognition problem in complex digital images using template matching. **International Journal of Natural Computing Research**, v. 1, n. 2, p. 54–70, 2010.

DATTA, R. et al. Image retrieval: Ideas, influences, and trends of the new age. **ACM Computing Surveys**, v. 40, n. 2, p. 5:1–5:60, 2008.

ER, M. J.; CHEN, W.; WU, S. High-speed face recognition based on discrete cosine transform and RBF neural networks. **IEEE Transactions on Neural Networks**, v. 16, n. 3, p. 679 –691, may 2005.

ETEMAD, K.; CHELLAPPA, R. Discriminant analysis for recognition of human face images. **Optical Society of America**, v. 14, n. 8, p. 1724–1733, 1997.

FAZL-ERSI, E.; TSOTSOS, J. Local feature analysis for robust face recognition. In: **Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications**. [S.l.: s.n.], 2009. p. 1–6.

FERNANDEZ, C.; VICENTE, M. Face recognition using multiple interest point detectors and SIFT descriptors. In: **Proceedings of the Eigth IEEE International Conference on Automatic Face Gesture Recognition**. [S.l.: s.n.], 2008. p. 1–7.

FINLAYSON, G. D.; CHATTERJEE, S. S.; FUNT, B. V. Color angular indexing. In: **Proceedings of the Fourth European Conference on Computer Vision**. [S.l.: s.n.], 1996. p. 16–27.

GENG, C.; JIANG, X. Face recognition based on the multi-scale local image structures. **Pattern Recognition**, v. 44, p. 2565–2575, 2011.

GONZALEZ, R.; WOODS, R. **Digital Image Processing**. 3rd. ed. [S.l.]: Prentice Hall, 2009.

GROSS, R.; BRAJOVIC, V. An image preprocessing algorithm for illumination invariant face recognition. In: **Proceedings of the Fourth International Conference on Audio- and Video-based Biometric Person Authentication**. [S.l.: s.n.], 2003. p. 1018.

GUO, H.; YU, Y.; JIA, Q. Face detection with abstract template. In: **Proceedings of the Third International Congress on Image and Signal Processing**. [S.l.: s.n.], 2010. v. 1, p. 129–134.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: **Proceedings of the fourth Alvey Vision Conference**. [S.l.: s.n.], 1988. v. 15, p. 147–151.

HSU, C.-Y. et al. Automatic extraction of face contours. In: **Proceedings of the 2010 International Joint Conference on Neural Networks**. [S.l.: s.n.], 2010. p. 1–8. ISSN 1098-7576.

HSU, R.-L.; ABDEL-MOTTALEB, M.; JAIN, A. K. Face detection in color images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 24, n. 5, p. 696–706, 2002.

KANAN, H. R.; FAEZ, K.; HOSSEINZADEH, M. Face recognition system using ant colony optimization-based selected features. In: **Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defense Applications**. [S.l.: s.n.], 2007. p. 57–62.

KARABOGA, D. **An Idea Based on Honey Bee Swarm for Numerical optimization**. [S.l.], 2005. Disponível em: <http://mf.erciyes.edu.tr/abc/>.

KARABOGA, D.; AKAY, B. A comparative study of artificial bee colony algorithm. **Applied Mathematics and Computation**, v. 214, n. 1, p. 108–132, 2009.

KARUNGARU, S.; FUKUMI, M.; AKAMATSU, N. Face recognition using genetic algorithm based template matching. In: **Proceedings of the IEEE International Symposium on Communications and Information Technology**. [S.l.: s.n.], 2004. v. 2, p. 1252–1257.

KENNEDY, J.; EBERHART, R. C. Particle swarm optimisation. In: **Proceedings of the IEEE International Conference on Neural Networks**. [S.l.: s.n.], 1995. p. 1942–1948.

KIM, T.-K. et al. Component-based LDA face description for image retrieval and MPEG-7 standardisation. **Image and Vision Computing**, v. 23, n. 2005, p. 631642, 2005.

KRIBY, M.; SIROVICH, L. Application of the Karhunen-Loeve procedure for the characterization of human faces. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 12, n. 1, p. 103–108, 1990.

LIHONG, Z. et al. Face recognition based on image transformation. **Global Congress on Intelligent Systems**, v. 4, p. 418–421, 2009.

LIN, C. Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network. **Pattern Recognition Letters**, v. 28, n. 16, p. 2190–2200, 2007.

LIN, C.-J. et al. 2D/3D face recognition using neural networks based on hybrid Taguchi-Particle Swarm Optimisation. In: **Proceedings of the Eighth International Conference on Intelligent Systems Design and Applications**. [S.l.: s.n.], 2008. p. 307–312.

LISIN, D. A. et al. Combining local and global image features for object class recognition. In: **IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2005. v. 03, p. 47–54.

LIU, Z.; LIU, C. A hybrid color and frequency features method for face recognition. **IEEE Transactions on Image Processing**, v. 17, n. 10, p. 1975–1980, 2008.

LIU, Z.; LIU, C. Fusion of color, local spatial and global frequency information for face recognition. **Pattern Recognition**, v. 43, p. 2882–2890, 2010.

LOWE, D. G. Object recognition from local scale-invariant features. In: **Proceedings of the International Conference on Compter Vision**. [S.l.: s.n.], 1999. p. 1150–1157.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 2, n. 60, p. 91–110, 2004.

MANJUNATH, B. et al. Color and texture descriptors. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 11, n. 6, p. 703–715, 2001.

MATSUMOTO, A. et al. Robust face recognition under various illumination conditions. **IEICE Transactions on Information & Systems**, E89-D, n. 7, p. 2157–2163, 2006.

MEHROTRA, H.; SA, P. K.; MAJHI, B. Fast segmentation and adaptive SURF descriptor for iris recognition. **Mathematical and Computer Modelling**, 2012.

MIKOLAJCZYK, K.; SCHMID, C. Scale & affine invariant interest point detectors. **International Journal of Computer Vision**, v. 60, n. 1, p. 63–86, 2004.

MIKOLAJCZYK, K.; SCHMID, C. A performance evaluation of local descriptors. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, n. 10, p. 1615 –1630, October 2005.

NABATCHIAN, A.; ABDEL-RAHEEM, E.; AHMADI, M. Illumination invariant feature extraction and mutual-information-based local matching for face recognition under illumination variation and occlusion. In: **International Jornal on Pattern Recognition**. [S.l.: s.n.], 2011. v. 44, n. 10-11, p. 2576–2587.

NETO, H. V. On-line visual novelty detection in autonomous mobile robots. In: ____. **Introduction to Mordern Robotics V.2**. Annerley, Australia: iConcept Press, 2011. v. 2, p. 241–265.

NIST. **Face Recognition Format for Data Interchange - Best Practices**. United States, 2007.

NSTC. **Face Recognition - Document of Subcomittee on Biometrics**. [S.l.], 2006. Disponível em: <www.biometrics.gov>.

OJALA, T.; PIETIKAINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on features distributions. **Pattern Recognition**, v. 29, n. 1, p. 5159, 1996.

PAWAR, V.; TALBAR, S. N. An investigation of significant object recognition techniques. **International Journal of Computer Science and Network Security**, v. 9, n. 5, p. 17–29, 2009.

PEREZ, C. B.; OLAGUE, G. Evolutionary learning of local descriptor operators for object recognition. In: **Proceedings of the Eleventh Annual Conference on Genetic and Evolutionary Computation**. [S.l.: s.n.], 2009. p. 1051–1058.

PERLIN, H. A.; LOPES, H. S.; CENTENO, T. M. Particle swarm optimization for object recognition in computer vision. **Lecture Notes in Computer Science**, v. 5027, p. 11–21, 2008.

PIMENOV, V. Fast image matching with visual attention and SURF descriptors. In: **Proceedings of the 19th International Conference on Computer Graphics and Vision**. [S.l.: s.n.], 2009. p. 49–56.

PRIYA, K. J.; R.S.RAJESH. Selective local features based face recognition with single sample per class. In: **Jornal of the Brazilian Computer Society**. [S.l.: s.n.], 2012. v. 18, p. 229–225.

PRODOSSIMO, F. das C.; CHIDAMBARAM, C.; LOPES, H. S. Proposta de uma metodologia para a construç ão de um banco de imagens faciais normalizadas. In: **Anais do VIII Workshop de Visão Computacional**. [S.l.: s.n.], 2012.

PRODOSSIMO, F. das C.; CHIDAMBARAM, C.; LOPES, H. S. Eye detection approaches based on adaboost and SVM algorithms. In: **Anais do IX Workshop de Visão Computacional**. [S.l.: s.n.], 2013.

RAHMAN, S. et al. Performance of MPEG-7 edge histogram descriptor in face recognition using Principal Component Analysis. In: **Proceedings of the Thirteenth International Conference on Computer and Information Technology**. [S.l.: s.n.], 2010. p. 476–481.

RAMADAN, R. M.; ABDEL-KADER, R. F. Face recognition using particle swarm optimization-based selected features. **International Journal of Signal Processing, Image Processing and Pattern Recognition**, v. 2, n. 2, p. 51–66, 2009.

SCHMID, C.; MOHR, R.; BAUCKHAGE, C. Evaluation of interest point detectors. **International Journal of Computer Vision**, v. 37, n. 2, p. 151–172, June 2000. ISSN 0920-5691.

SHAN, S. et al. Extended fisherface for face recognition from a single example image per person. In: **Proceedings of the IEEE International Symposium on Circuits and Systems**. [S.l.: s.n.], 2002. v. 2, p. 81–84.

SINGH, C.; WALIA, E.; MITTAL, N. Robust two-stage face recognition approach using global and local features. **The Visual Computer**, v. 28, n. 11, p. 1085–1098, 2012.

SOLAR, J. R. del; QUINTEROS, J. **Comparing pre-processing approaches for illumination invariant face recognition**. [S.l.], 2006. Disponível em: <http://vision.die.uchile.cl>.

SOLAR, J. R. del; QUINTEROS, J. Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches. **Pattern Recognition Letters**, v. 29, n. 14, p. 1966 –1979, 2008.

STRUC, V.; PAVESIC, N. A case study on appearance based feature extraction techniques and their susceptibility to image degradations for the task of face recognition. **World Academy of Science, Engineering and Technology**, v. 54, p. 811–819, 2009.

SUGISAKA, M.; FAN, X. An effective search method for neural network based face detection using particle swarm optimization. **IEICE Transactions on Information & Systems**, E88-D, n. 2, p. 214–222, 2005.

TAN, X. et al. Face recognition from a single image per person: A survey. **Pattern Recognition**, v. 39, p. 1725–1745, 2006.

TERESHKO, V.; LOENGAROV, A. Collective decision-making in honey bee foraging dynamics. **Computing and Information Systems**, v. 9, n. 3, p. 1–7, 2005.

TOLBA, A. S.; EL-BAZ, A. H.; EL-HARBY, A. A. Face recognition: A literature review. In: **International Journal of Information and Communication Engineering 2:2**. [S.l.: s.n.], 2006. p. 88–103.

TRUJILLO, L.; OLAGUE, G. Synthesis of interest point detectors through genetic programming. In: **Proceedings of the Genetic and Evolutionary Computation Conference**. [S.l.: s.n.], 2006. p. 887–894.

TRUJILLO, L.; OLAGUE, G. Automated design of image operators that detect interest points. **Evolutionary Computing**, v. 16, n. 4, p. 483–507, December 2008.

TURK, M. A.; PENTLAND, A. P. Face recognition using eigenfaces. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 1991. p. 586–591.

VAN de WEIJER, J.; SCHMID, C. Blur robust and color constant image description. In: **IEEE International Conference on Image Processing**. Piscataway, USA: IEEE Press, 2006. p. 993 –996.

VENKATESAN, S.; MADANE, S. S. R. Face recognition system with genetic algorithm and ANT colony optimisation. **International Journal of Innovation, Management and Technology**, v. 1, n. 5, p. 469–471, 2010.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2001. v. 1, p. 511–518.

WANG, J.; YANG, H. Face detection based on template matching and 2DPCA algorithm. In: **Proceedings of the Congress on Image and Signal Processing**. [S.l.: s.n.], 2008. p. 575–579.

WISKOTT, L. et al. Face recognition by elastic bunch graph matching. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 7, p. 775–779, July 1997.

YANG, J. et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 26, n. 1, p. 13 –137, January 2004.

YANG, M.-H.; KRIEGMAN, D. J.; AHUJA, N. Detecting faces in images: A survey. **IEEE Transactions on Pattern Recognition and Machine Intelligence**, v. 21, n. 1, p. 34–58, 2002.

ZHANG, D.; CHEN, S.; ZHOU, Z.-H. A new face recognition method based on SVD perturbation for single example image per person. **Applied Mathematics and Computation**, v. 163, n. 2, p. 895–907, 2005.

ZHAO, W. et al. Face recognition: A literature survey. **ACM Computing Surveys**, v. 35, n. 4, p. 399–458, 2003.

ZHOU, D. et al. Improved-LDA based face recognition using both facial global and local information. In: **Pattern Recognition**. [S.l.: s.n.], 2006. v. 27, p. 536–543.

ZHOU, H.; SCHAEFER, G. Semantic features for face recognition. In: **Proceedings of the Fifty second International Symposium ELMAR -2010**. [S.l.: s.n.], 2010. p. 33–36.

ZHOU, W.; AHRARY, A.; KAMATA, S.-I. Image description with local patterns: An application to face recognition. **IEICE Transactions on Information & Systems**, E95-D, n. 5, p. 1494–1505, 2012.

ZULIANI, M.; M.KENEDY; MANJUNATH, B. A mathematical comparison of point detectors. In: **Proceedings of the Computer Vision and Pattern Recognition Workshop**. [S.l.: s.n.], 2004. p. 172–197.

APÊNDICE A – STILL IMAGES WITH MULTIPLE FACES

(a)

(b)

(c)

Figura 31: Some sample images under illumination Conditions (a) IL-I (b) IL-II (c) IL-III

(a)

(b)

(c)

Figura 32: Some sample images under illumination Conditions (a) IL-I (b) IL-II (c) IL-III

(a)

(b)

(c)

Figura 33: Some sample images with rotation

(a)

(b)

(c)

Figura 34: Some sample images with occlusion