University of Alberta

Face Recognition for Biometrics-Based Personal Identification

by

Gabriel Jarillo Alvarado ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

Edmonton, Alberta
Spring 2007

Your file Votre référence
ISBN: 978-0-494-29694-3
Our file Notre référence
ISBN: 978-0-494-29694-3

To God, my parents, Mayra, and friends
who encouraged me throughout my graduate studies.

Abstract

A thorough investigation in the area of face recognition is presented. We contemplate a variety of methods for the construction of feature spaces in face recognition. Linear and non-linear methods such as Eigenfaces, Fisherfaces, Isomap, and kernel-PCA are evaluated in terms of classification performance and robustness. The experimental environment comprehends well-known standards in the area of face recognition, namely YALE and FERET databases. Various lines of research are established in order to reveal the conditions in which the classifiers are likely to fail or thrive. Our investigations include:

- *The assessment of face classifiers in the presence of environmental disturbances.* We include models of deterioration of visual information that mimic frequent scenarios in face recognition.
- *The evaluation of a modular approach to face recognition.* We deliver an investigation on a modular approach within the framework of Principal Component Analysis (PCA). We comment on classifier performance and computational implications.
- *The study of the impact of image quality in face classifiers.* We report on an extensive investigation aimed at revealing the relationship between classifiers performance vis-à-vis anticipated levels of image resolutions. To our knowledge this is the first investigation that quantifies the behavior of linear and non-linear face classifiers with respect to image quality. Useful design recommendations are drawn from this investigation.
- *The assessment of aggregation of classifiers based on image transformations.* This is the first investigation putting together an assessment of numerous face classifiers and combined experts in the presence of image transformations. This is, considering both -

linear and non-linear methods for constructing feature spaces. Descriptors are constructed from using contrast enhancement and edge detection. Useful findings are discussed.

- *The exploration of an evolutionary approach towards improving classifier performance.* In this investigation we reveal in quantitative terms the importance of the features in a given face space. We present evidence of improvements in classification.

We offer useful design guidelines and recommendations for the architectures under investigation. We comment on identified advantages and drawbacks of each architecture based on the experimental findings.

Acknowledgements

I would like to express my most sincere appreciation to Dr. Witold Pedrycz and Dr. Marek Reformat for their excellent advice and encouragement throughout my graduate studies. I am grateful for their priceless guidance towards my academic and professional development.

The research project used the *FERET* database of facial images collected under the FERET program. The use of the Yale dataset is also acknowledged. Support from the Natural Sciences and Engineering Research Council (NSERC) and Canada Research Chair (Prof. W. Pedrycz) is gratefully acknowledged. Support from Consejo Nacional de Ciencia y Tecnología – CONACyT (The Mexican National Council for Science and Technology) is also appreciated.

Table of Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Motivation

This research is directly motivated by the necessity of personal identification methods that demonstrate reliability and robustness against fraud. In modern life, authorization and authentication of individuals have become vital components of day-to-day activities. Let it be for making financial transactions, driving a car, or crossing borders, we require truthful methods to prove our identities. While it is important to accurately identify a person, it is also imperative to prevent an identity from being misused or stolen. As of today, most methods for identity verification rely on "things" we know or possess, for instance passwords or identification cards. Since identification documents may be lost, forged or stolen and passwords may be forgotten or compromised, the traditional methods for personal identification are merely insufficient. Evidently the use of passwords or identity documents does little to prevent so-called "identity theft" - a problem that has been increasing over the last years, and that brings about negative economical implications. As examples, according to the 2003 report by Sinovate for the Federal Trade Commission (FTC) [112], the economical cost of identity theft in the U.S. was approximately $5 billion at that time. The total losses suffered by individuals and businesses worldwide are estimated at $221 billion during the same year [37]. The call for new and reliable methods for personal identification is clear.

An emerging technology that offers solutions to the current limitations of personal identification is based on biometrics. In a nutshell, biometrics-features are measurements or "metrics" taken from an individual based on his/her physiological or behavioral characteristics. The term "biometrics" is generally used to refer to the technology behind these metrics themselves. Biometric features are universal, unique, collectable, permanent, and non-transferable. Some representative examples include fingerprints, faces, or handwriting. By nature, biometric features are extremely difficult to reproduce or imitate, hence the potential of reducing identity misuse. The advantages of biometrics ensure a vital future for civilian and military applications. Biometrics technology can be useful for granting access to resources, authentication at ATM machines, identifying a particular individual in surveillance imagery, or automating the search for individuals in large databases, e.g. police records. While biometrics technology surely offers improvements over current identification methods, a single biometric may not be sufficient to develop a reliable system. As a matter of fact a robust identification system would probably require a combination of biometrics and other methods of personal identification.

The fact that we are naturally guided to identifying individuals through visual information gives face recognition a wide range of applications, thus it becomes an appealing biometric to investigate. Broadly speaking, the idea behind the face recognition process is to capture the distinctiveness of a face without being overly sensitive to noise, especially to lighting conditions, environmental disturbances, and facial expressions. Approaches to face recognition can be divided into two categories [12]:

- *Face appearance.* The underlying idea is to reduce the dimensionality of the original space (depicting pixel values) to a handful of numbers obtained by some transformation.
- *Face geometry.* The geometry of facial features such as eyes, nose, and mouth, are used to describe a person.

Over the years, methods embracing the notion of face appearance have demonstrated superior performance than those based on face geometry; therefore we report on face recognition techniques falling under the category of face appearance. The experimental environment concerns a well-known standard in this area such as the FERET [81] and YALE [120] databases. Both of them contemplate realistic scenarios in terms of lighting conditions and facial expressions. The images depict frontal views, showing only the face area, and the images are already aligned and centered.

There are two major scenarios in which face recognition, or any biometric system, can operate:

- *Individual identification.* The system attempts to single out the identity of a particular individual based on his/her biometric, answering the question: who is that person? The classification performance of the system is commonly expressed in terms of error rates.
- *Intruder detection.* The system attempts to accept or reject an individual from a group of known individuals, answering the questions: do we know that person? and, is he/she an intruder? The classification performance is expressed in terms of False Rejection Rates (FRR) and False Acceptance Rates (FAR). FRR indicates the chances of rejecting a genuine individual and FAR refers to the probability of accepting an intruder.

It is clear that if a system can identify any particular individual successfully, then both scenarios are covered. For such reasons we concentrate on individual identification throughout this research.

Interestingly enough, humans can identify faces with relatively little or no effort at all. However, building an automated system that accomplishes such tasks is a challenging problem. Examples of current commercial systems include:

- *FaceVACS* by Cognitec Systems [113]. The system can operate in environments where access is restricted or where the identity of individuals is to be found within a collected database. The system exhibits an identification rate of around 72% with large databases of still images.
- *FaceEnforce* by Cybula [114]. This is s a stand-alone system implemented in C++ mainly targeted at Sun platforms. It can search for individuals in large databases.
- *MIRH eye surveillance* by DreamMirh Co., Ltd. [115]. The system finds a particular individual in environments like airports, docks, and building lobbies. It requires a surveillance infrastructure in place to collect the data.

- *FaceIt ARGUS Screening System* by Identix Incorporated [116]. The system captures faces in live video and acts like a filter, searching against watchlist databases to see if a person may be wanted.
- *VeriLook* by Neurotechnologija [117]. This is a face localization and detection system capable of processing multiple faces in live video streams and still images. The system can identify individuals within a collected database.
- *Affinity Face Recognition* by OmniPerception [118]. It features face detection and identification from still images.

Unfortunately, most vendors of commercial systems do not disclose essential technical details regarding the underlying algorithms. They are also silent on the performance of their products. It is very difficult to compile a fair assessment of commercial systems and our face recognition methods.

As of today face recognition technology is still impractical in many realistic scenarios. Computational cost and classification accuracy are major obstacles to overcome in order to take advantage of this biometrics. Aspects such as size of databases, lighting conditions, and environmental disturbances could exhibit a negative effect on recognition rates, which lead us to concrete objectives for this research.

1.2 Objectives

The objectives of this study, falling under the umbrella of pattern recognition, are to investigate current and new approaches to face recognition, to identify strengths and weaknesses of different methodologies, and to propose design guidelines and new systems that cope with current limitations. There are well-identified factors that influence face classification that demand investigation. Clearly defined objectives are:

- *Assessment of linear and non-linear methods for dimensionality reduction.* At this point it is constructive to put together an evaluation of useful methods in terms of their computational costs, classification rates, and the conditions at which the classifiers are most likely to fail and thrive, e.g. lighting conditions, particular facial gestures, and size of databases.
- *Impact of image quality on performance of face recognition.* The study includes deteriorated and enhanced images. Degradation of visual information is perhaps the most frequent and damaging factor in face classification. We include frequent types of image degradation in our experimentation. A careful and methodical investigation is crucial to understanding the limits of current algorithms in realistic scenarios. On the other hand, improving the quality of images, particularly in terms of lighting conditions, proves valuable for correct classification. Meticulous examination of this area is required to develop sound and valid design recommendations.
- *Effect of image transformation in classification (e.g. edge detection).* It is known that edge detection considerably overcomes problems of lighting conditions when it comes to recognizing shapes. In the case of face recognition, edge detection can provide more distinctive information regarding the outlines of the face and facial features of each individual. Our edge detection process was improved by a pre-processing step of contrast enhancement.
- *Influence of image resolution on performance of classifiers.* High computational cost commonly leads to the usage of low quality images in face recognition. However, there

is not a clear understanding as to what is the minimum or adequate resolution to consider while preserving system accuracy. As of today, there isn't a comprehensive study that takes such tradeoffs into consideration. We present a thorough investigation that reveals the performance of face classifiers under various image resolutions.

- *Aggregation of classifiers combining distinct features.* In this regard we investigate in the collective knowledge of different features towards robust and accurate classification. The combination of different features can lead to better descriptions of individuals, hence revealing relevant distinctive information between persons. The features taken into consideration were computed from original images and images transformed by histogram equalization and edge detection.
- *Investigation on evolutionary optimization in face recognition.* This is an effort to improve classification quality (reducing classification error) by finding the importance, from the classification point of view, of each variable (feature) in a given feature space. We also seek for the most suitable similarity measure for distinguishing individuals more effectively.

1.3 Contributions

Bearing the outlined objectives in mind, the findings of our research activities offer valuable contributions to the field of face recognition. We report on the performance of the classifiers, which is based on a comprehensive suite of experiments, and deliver several design hints supporting further developments of face classifiers. Key contributions are enumerated:

1. *The thorough evaluation of the performance of two common face classifiers.* At this point we evaluate the performance of Eigenfaces [98] and Fisherfaces [10] operating in the presence of deterioration of available visual information. The findings of our study are crucial to identify at which levels of noise the face classifiers can still be considered valid. Prior knowledge helps develop adequate face recognition systems. We investigate several typical models of image distortion such as Gaussian noise, salt and pepper, and blurring effects and demonstrate their impact on the performance of the classifiers. Several distance models derived from the Minkowski family of distances are investigated with respect to the produced classification rates. We offer design guidelines and useful recommendations towards improving recognition performance.
2. *The assessment of the effects of image resolution and image transformations in face recognition.* Our findings portray practical implications to systems design. Image transformations include contrast enhancement via histogram equalization, and edge detection by means of the Sobel operator. The methods for dimensionality reduction involve Eigenfaces, Fisherfaces, Isomap [95], and kernel-PCA [90]. Through extensive experimentation, we reveal and quantify the tradeoff between image resolution and performance of the classifiers. The fact that this study takes into consideration several methods for dimensionality reduction, image transformations, and image resolutions makes this contribution unique.
3. *The set of recommendations regarding certain aggregation of features emerging from regular images, edge images, and histogram-equalized images.* The features are computed using several methods for dimensionality reduction that prove useful, such as Eigenfaces, Fisherfaces, Isomap, and Kernel-PCA. The findings of our exploration on impact of image resolution helped us decide on an appropriate image size without

loosing accuracy, yet reducing computational cost. We offer evidence of classification improvement over traditional methods as well as design recommendations.

4. *The approach suitable for identifying significant features for classification purposes.* Most face classifiers take into consideration certain features (variables) represented in a given feature space to achieve most favorable recognition rates. The features are commonly selected by ranking them according to their particular variances. Nonetheless the variances may not reflect the true "importance" of such variables towards correct classification. We put forward a method capable of revealing the importance of each variable from the classification point of view. Our method also finds a suitable similarity measure to distinguish individuals in a given feature space. Evolutionary optimization by Genetic Algorithms [100] ranks each variable and produces a suitable distance model in order to separate individuals. Evolution is driven towards reducing error rates. Experimental evidence supports the usefulness of this approach in various scenarios. We also comment on major advantages and potential limitations of the proposed architecture.

Chapter 2

BIOMETRICS – AN INTRODUCTION

This section presents a general description of the diversity of biometric technologies, and places face recognition as one of the most predominant and appealing areas for research. The introduced literature review intends to provide the reader with a general background of the most predominant trends in face recognition technologies. The existing literature on face recognition is extensive. This section only covers what we consider to be the most illustrative part of it, allowing us to grasp a general overview of the face recognition research in terms of its diversity, trends, and achievements. It also discusses the progress of face recognition systems and their limitations. The initial paragraphs of this chapter describe the larger field of biometrics, always keeping special interest in face recognition and showing where it fits in. The latter paragraphs provide a general description of face recognition itself.

Biometrics is a relatively new field that seeks the correct and robust identification and verification of individuals. Although biometrics emerged from its use in law enforcement to identify criminals, it is increasingly being used to establish person recognition in a large number of civilian applications [34].

The notion behind biometrics departs from the fact that every individual is different in nature, therefore there are unique characteristics that make that person distinguishable from others. This concept is fundamental for the existence of biometric-based systems, and allows us to outline some requirements for the persons' characteristics to qualify as valid biometric features [34]:

- *Universality*: each and every person should have the characteristic.
- *Distinctiveness*: Any two persons should be sufficiently different in terms of the characteristic.
- *Permanence*: the characteristic should be sufficiently invariant (with respect to the matching criterion) over a valid period of time.
- *Collectability*: the characteristic can be measured quantitatively.

There are several definitions of biometrics in the literature, one that embraces the concept in a concise manner reads as follows: "Biometrics are automated methods of recognizing a person based on his/her physiological or behavioral characteristic [119]". It is known that every person is distinguishable from others, even twins, from the physiological and behavioral points of view. By embracing such a fact we can organize a biometric as belonging to one of two possible categories, either physiological or behavioral. Each person has his/her own looks, personality, way to react to situations, way to behave, etc. making him/her an "individual" within the full context of the word. Ensuing the previous rationale one can provide examples of common features to use in biometrics. We list some of them by category, either physiological or behavioral:

2.1 Biometrics based on physiological characteristics

Face: It is probably the most common non-intrusive biometric used for personal recognition. Many popular approaches take into consideration either the location and shape of facial attributes (i.e. eyes, mouth, nose), or the entire face. The latter is also referred to as global analysis of face images. It represents a face as a weighted combination of a number of canonical faces [36].

Fingerprints: it is a very reliable non-intrusive method for personal identification. The fingerprint is the pattern of ridges and valleys on the surface of the fingertip. They are so unique that even the fingerprints of twins are different [36]. In [67], Maio, *et al.* has shown an extensive analysis of a number of algorithms applied in this area that exhibit encouraging results.

Hand geometry: these systems are based on a number of measurements taken from the human hand, including its shape, palm size, and lengths and widths of the fingers. However the geometry of the hand is not very distinctive. In addition, systems based on this technology cannot be scaled up for identification of individuals in large populations [36].

Iris: the iris is the colored part of the eye surrounding the pupil, located behind the cornea and the aqueous humour. Development of the iris begins by the 3rd month of gestation and it is already highly developed by the age of 12 months, after that period it remains stable for life. Each iris is unique, even in identical twins. The probability of 2 irises producing the same IrisCode is approximately 1 in 10^{78} [87]. Some characteristics include freckles, pits, striations, and vasculature. An Iris Scanning system can supposedly detect a live eye by virtue of the small continuous fluctuations in the pupil, and by the natural physiological response of the pupil to light [87].

Voice: speaker recognition systems can be divided into two categories, namely text-dependent, and text-independent. In text-dependent systems, the user is expected to use the same text during training and recognition sessions [99]. A text independent system does not use the training text during recognition session. The speech signals corresponding to a test phrase of a group of people are recorded in voice files on a computer. The information is then converted from the time domain to the frequency domain using digital signal processing techniques [99].

2.2 Biometrics based on behavioral characteristics

Handwriting: these methods extract features such as writing speed, direction, duration, height, width, slant angle, black pixels etc. They rely on the interactive localization and segmentation of the relevant text information. Since the purpose this technology is to identify the writer of a specific handwriting, the recognition system is not concerned about the content of the written text [91].

Keystroke: typing biometrics is defined as the analysis of a user's keystroke patterns. Each person has an almost unique pattern of typing. This pattern can be learned with the purpose of identifying a particular user [30].

Gait: it refers to the particular way a person walks in a complex spatio-temporal context. Gait is not very distinctive, but it is sufficiently discriminatory to allow verification in some low-security applications [36].

As a matter of fact, the list of potential biometrics for personal identification is continuously growing. As the science advances it is possible to find new features suitable for identification purposes.

Overall, biometrics promises to tackle the problems that affect traditional methods for personal identification. As of today, current methods of identification involve possession of tokens, such as passports or driver's licenses; they also include knowing passwords, such as Personal Identification Numbers or Social Security Numbers. In either case they can be counterfeited, stolen, forgotten or compromised. Unlike tokens or passwords, biometric identifiers are inextricably linked to the persons themselves, thus making the technology much more difficult to breach [121]. This is a strong argument in favour of biometrics-based identification systems that leaves most of the current technologies obsolete.

As the level of security violations and fraudulent activities increase, the need for highly secure identification and personal verification technologies becomes evident. Biometric-based solutions aim at providing secure instruments for confidential transactions and personal data privacy. The need for biometrics can be found in the federal and local governments, in the military, and in commercial and civilian sectors [119]. In fact the practical applications for biometric-based systems are extensive. However the most common ones are related to security and law enforcement [50]. Typical environments where biometrics are convenient include: giving access to resources, correct implementation of security services, and facilitating border crossing. The selection of the most appropriate biometrics to use depends on the particular needs and requirements of the practical application, as examples one may consider attributes such as costs of the technology, effort to use it and implement it (this includes maintenance), intrusiveness to the users, and accuracy of the overall system. A useful comparative analysis for choosing the most suitable biometrics is provided by the Zephyr charts. By deciding on the overall criteria for selection, one can eliminate the biometrics that do not qualify as best. Figure 1 (illustration taken from [12]) presents an example of the Zephyr chart in which a biometric-feature is chosen from several candidates based upon their common attributes. In our example the attributes include effortlessness, non-intrusiveness, inexpensiveness, and accuracy. The attributes are defined according to specific requirements in a particular environment. Based upon specified thresholds on the area A, one can eliminate those biometrics that are inadequate. The larger the area A, the better the biometric [12]. In the example depicted in Figure 1, iris technology is suggested as the best.

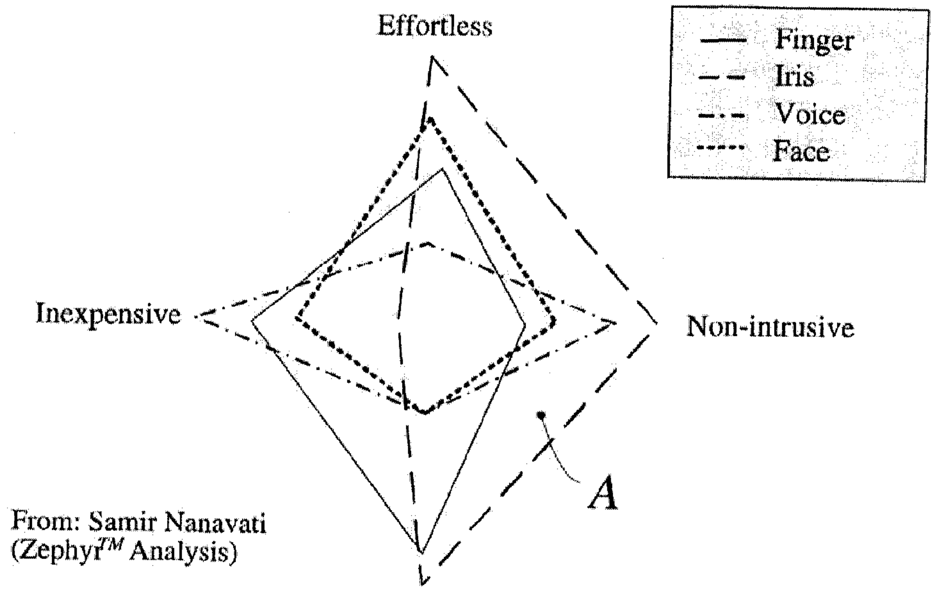


Figure 1 Example of Zephyr chart

As a matter of fact a single biometric may not be sufficient to correctly identify an individual. A number of them in combination with other methods of identification may be necessary to design a reliable system. Due to the extensiveness of this field of study we focus our attention to face recognition alone.

2.3 Face recognition – An overview

The fact that we are naturally guided to use face images for identification purposes makes face recognition an appealing biometric to investigate. Facial images are already widely used and generally accepted for personal identification, and as such, they are used in a large variety of official documents, for example passports, drivers' licenses, and some credit cards. So far the task of matching faces ultimately relies upon people, hence it is constrained by the limited number of faces a person can process in relatively short period of time. Imagine finding few faces from thousands of face images, for example in police records; or how can a security institution identify someone wanted when he/she is walking in any public place? What about identifying someone known for car theft in a parking lot? There are already surveillance cameras around major cities in public places, government buildings, businesses, schools, ATMs, etc. Surveillance systems are used to monitor traffic, public gatherings and sometimes to identify suspects. Face recognition is a technology that portrays enormous importance in systems that require automated search of individuals, saving time and money in the long run. We envision that if the face recognition process becomes reliable and automated, the spectrum of practical applications for this technology is endless, from granting access to PCs at home to border crossing and homeland security.

2.4 Problem definition

It is quite interesting that even though humans can detect and identify faces with relatively little effort, building an automated system able to accomplish such tasks is still very challenging [62]. The enormity of the problem has involved hundreds of scientists and interdisciplinary researchers, and yet it is still difficult to design an automated system capable of fully addressing the intricacies of this task [24]. Researchers in psychology, neural sciences, engineering, image processing, and computer vision have investigated a number of issues related to the process of face recognition by humans and machines. Various physical aspects such as hairstyle, face orientation, gesture, non-uniformity of the background, long-range surveillance footage, variations of lighting, image quality, and noisy environments affect recognition performance. For the latter, it is common to encounter blurred images due to misfocus of lenses, noise introduced by the image acquisition system, and other distortions intrinsic to the image processing methods. Overall, the face recognition process should be capable of capturing the distinctiveness of a face without being overly sensitive to noise [12].

In addition to the challenges outlined above, there are several internal issues that affect the face recognition process. For example, the dimensionality of the image space (images depicted by pixels) is quite broad. The extraction and construction of meaningful features is a must. Another issue that commonly arises in face recognition is the "Small Sample Size problem" (SSS) [65][66], this is, the number of available samples is far smaller than the dimensionality of the samples - a situation that leads to a poor representation of a class in the given space. In face recognition it is common to have databases comprising a large number of individuals, and yet just a few images for each person. The intrinsic consequences are a poor representation of each individual in a given feature space and low classification rates.

In general, face recognition of still images can be divided into two categories: geometric matching and template matching. In case of geometric matching, the geometric characteristics of faces are compared. For template matching, an image feature is extracted from an array of pixels (face image), and compared to a set of feature templates stored in a database. Many of the template matching approaches use principal component analysis as a vehicle that leads to the development of highly representative image features while reducing the dimensionality of the original face images. Usually a face image is projected onto a feature space spanned by some basis image functions, just like a Fourier transform projects an image onto basis images of the fundamental frequencies [12].

The face recognition process involves three main phases:

- *Detection*: advocated to finding all the faces depicted in an image.
- *Extraction*: Assigned to pulling out the information concerning faces.
- *Classification*: designed to match the extracted face information (features) with the cases (individuals) stored in the databases.

These three steps rely on a number of different methods from the areas of statistics, data mining, pattern recognition, and Computational Intelligence.

At the present time, there are some techniques for description and classification of human faces. However none of them successfully addresses the realistic requirements of dealing with large databases and noisy environments. Examples of some existing techniques are contour modeling, Eigenfaces, Fisherfaces, local feature analysis, and neural networks.

Despite the best efforts of countless researchers, the classification rates reflect poor performance when the systems deal with large databases. Typical numbers produced by these methods are in the range of 10% to 20% for intruder detection, and within 10% to 30% for personal identification. The results presented in the literature call for the improvement of the face recognition technology, especially for practical applications in realistic scenarios.

The problem is not simple yet the benefits are appealing. Our research admits limitations and sets some constrains in order to make it feasible. We do not undertake the task of finding and extracting face images from the pictures. We rather concentrate on classifying the faces already extracted, properly aligned, oriented, and stored in the databases. The face recognition problem as is, still confronts several major challenges to its success.

As of today there is no face recognition system that can successfully tackle the problems mentioned above. Face recognition is yet an unsolved problem that requires further consideration and ingenuity. As we know, humans have a remarkable ability to process face images. However, the process has not been understood completely so it is impossible to try to mimic it with a computational model [76]. Several attempts have been made in face recognition using statistical, evolutionary, data mining, and machine learning approaches, as outlined in the following literature survey.

Taking into consideration the issues described above, we envision that this research will contribute to the areas of pattern recognition and human face classification in terms of new methodologies and ensuing algorithms. The face recognition problem is complex yet very appealing. Our ultimate goal is to investigate on existing methods and to improve the reliability of biometric technology for practical applications. In this research we offer a fair and careful comparison and analysis of some face recognition algorithms, setting common ground for future research and development in the field. Based on this work, we provide useful guidelines and design recommendations for new face recognition systems. We envision that the contributions will be relevant for the construction of systems that are more reliable and efficient in face classification.

2.5 Approaches to face recognition – Literature survey

The existing literature concerning face recognition is extensive. We can track back the first formal method for classifying faces introduced by Galton in 1888, according to Barrett [8]. Galton proposed collecting facial profiles as curves, finding their norm, and then classifying other profiles by their derivations from the norm [8]. Since then, the face recognition topic has been of interest for the scientific community, nevertheless it is still somewhat elusive and challenging.

In 1973 Kanade [41] published in his doctoral thesis, extensive work in face recognition. He based his approaches mostly on geometrical characterization of facial features, such as mouth, eyes, and nose. He used distance measurements to find the appropriate classes among several individuals in a given database. In his work, Kanade used a database consisting of 20 persons, making the task rather challenging at that time.

In 1990, Kirby and Sirovich [46] proposed a statistical method for face description. Their motivation was focused mostly towards image compression rather than image classification. However, their findings are important for face recognition as well. They adopted the Karhunen-Loève expansion, commonly known as principal component analysis, to eliminate redundant information from facial images. They argued that since faces are symmetric to some extent, the information retained in the image pixels is redundant, and therefore reducible. Later in 1991, Turk and Pentland [98] popularized this technique in their famous paper "Face recognition using Eigenfaces" with direct applications to face recognition. This technique based on principal component analysis is now commonly known as "Eigenfaces".

The Eigenfaces method is widely used today, and has been adopted in most face recognition systems. Therefore dedicating a few lines for a brief description is worthwhile. In a nutshell, Eigenfaces makes use of the covariance matrix formed by given face images in order to "draw" similarities among their pixel values. It then extracts linearly independent data by means of principal components. The feature vectors are then constructed as weighted combinations of basis vectors obtained by PCA. The formal description of Eigenfaces is presented later in section 4.2. The Eigenfaces method has achieved popularity among the scientific community. However, it is known for having difficulties dealing with changes in facial expression and lighting - scenarios that are frequent in practical applications. On this issue Adini *et al.* [1] did a study in 1997 to quantify the variations attributed to changes in illumination direction between processed images, and compared them to variations between different individuals taken under the same illumination. They found out that variations due to illumination are almost larger than variations due to change in face identity. A good example of such a scenario is presented in Figure 2. The illustration is taken from [10]. Such problem exhorts for improvements in the algorithms and conception of new ideas.

Figure 2 Same individual under different lighting conditions

In 1997 Belhumeur, *et al.* [10] substantially tackled the problem of illumination. They proposed a statistical approach for face classification based on Fisher's Linear Discriminant (FLD) [25], hence the name of their method "Fisherfaces". In a few words and superficially speaking, Fisherfaces maximizes the ratio of the between-class and the within-class scatters of features projected in the space, making the classification task more suitable. Normally Fisherfaces produces well-separated classes in a low-dimensional subspace, even under large variations of facial gestures and lighting conditions. In their work, Belhumeur, *et al.* demonstrated the ability of Fisherfaces to deal more adequately with variations of lighting and facial expressions. They compared the popular Eigenfaces against Fisherfaces and outlined the advantages of their method. At present, Fisherfaces has become a valuable and valid technique for feature extraction and classification of human faces. Fisherfaces has been widely adopted by the scientific community in face recognition. For some examples refer to [60][61][15][19].

Eigenfaces and Fisherfaces have become the common platforms for feature extraction in face datasets. Some other features and techniques may be combined and applied afterwards to improve classification results. However, the presence of Eigenfaces and Fisherfaces in face recognition systems is colossal. Now that we have established the importance and omnipresence of Eigenfaces and Fisherfaces, we would like to dedicate some lines to briefly summarize and describe the different approaches and techniques suggested by other researchers in the field. As final words for this paragraph we would like to acknowledge the adoption of such methods as starting platforms for our research.

Along the same line of PCA, Kim, *et al.* [45] proposed a second order mixture of Eigenfaces for face recognition. In their work they obtained a set of vectors by transforming images using PCA. Then they obtained second-order Eigenfaces by applying PCA once more to the set of the residual vectors. The residual vectors are obtained by computing the difference between the original face images and the reconstructed images via the Eigenfaces method. They claim that their method overcomes the problems of variations in pose and illumination intrinsic to Eigenfaces. They tested their method using an MPEG-7 face dataset that contains 271 individuals with 5 different images of each person. Among those images, there are 740 images of 148 persons, and a pose-invariant data that comprises 615 images of 123 individuals. The images are collected from AR (Purdue), AT&T, Yale, UMIST, University of Berne, and from MPEG-7 news videos. The performance of their method was measured in terms of false identification rate, which they reported in the range 15-36%.

Some researchers have also suggested different features for face classification, for instance Hong [31] proposed using Singular Values (SV) obtained from the images in the training set as features for face recognition. He claimed that the singular values are very stable and insensitive to noise to an admitted extent. Another appealing characteristic is that singular values are invariant to algebraic and geometric transformation, such as rotation. Despite the contribution by Hong, later in 2002 Tian, *et al.* [96] found that in fact singular values do not contain enough information to adequately describe an image. Tian also stated that most of the important discriminatory information is stored in the two orthogonal matrices delivered by Singular Value Decomposition (SVD).

Reisfeld and Yeshurun [82] proposed another interesting approach in 1992. They used a generalized symmetry operator with the intention of finding the eyes and mouth in a face image. The symmetry operator produces a symmetry map that is used to assign a symmetry magnitude and a symmetry orientation to each point. They claim to have achieved a 95% success rate in their own image dataset.

In 1997 Tistarelli and Grosso [97] offered a face recognition system inspired in the way humans perceive scenes. They claim that the human vision system collects several pictures of the face by directing the gaze towards different points, such as eyes, lips, and nose. This mechanism limits the bandwidth of the signal to be processed, thus reducing the amount of time required for recognition. Their work mimics the concept of human vision by using space variant sampling of the images, giving higher resolution or "interest" to key elements in an image. After the task is completed they constructed a vector with the extracted information from every image. Later they applied the concept of Eigenfaces and performed classification via the nearest neighbor classification rule. When their method presented discrepancies in the suggested class they used histograms of the gray level values to alleviate the classification outcome. They report a recognition rate of 98% on a dataset containing 152 images of 19 subjects.

In the area of neural networks for face recognition, Aitkenhead, *et al.* [2] implemented two cascaded Neural Networks (NN). The first NN detects the presence of a face in an image using edge detection methods as part of the feature extraction, and the second NN is entrusted the task of providing the appropriate class. The input features of the second NN are binary images depicting the edges of the objects in the face (i.e. eyes, nose, and mouth). As for the face detection NN, the input features are trimmed images of 32x32 pixels depicting only the face area (in which manual intervention is required to align the images). Their database comprises images extracted from video sequences, and include 5 images per individual for a total of 20 persons. The images include variations in the background as well as image rotations (-30^0, -15^0, 0^0, 15^0, 30^0). However, it does not include variations in lighting. The authors reported 94.7% correct face detection, and 75% correct face classification.

In the context of geometric features for face recognition and classification, Lin, *et al.* [58] proposed a "spatially eigen-weighted Hausdorff distances for human face recognition". In their work they obtained the average image of the training set. This average image is a binary representation that only shows the edges of the facial features (i.e. nose and mouth). They applied a number of processes in order to obtain the binary images, such as filtering to emphasize the edges, and thresholding to produce the binary values. They formed a correlation matrix using the average image and the images in the dataset, somewhat resembling a covariance matrix. They performed PCA over the correlation matrix. From the set of eigenvectors and eigenvalues delivered by PCA, they selected the one with the highest variance and propose it as a "mask" or "weight vector" for the Hausdorff distance. The classification task takes place in the form of a nearest neighbor classifier. They reported a classification rate of about 83%.

Chapter 3

FACE RECOGNITION – METHODOLOGY AND RESEARCH ENVIRONMENT

This chapter provides a description of the research methodology and the research environments adopted in our investigations. They are carefully explained in the section labeled as "experimental scheme". The experimental scheme takes into consideration various methods for constructing some meaningful feature spaces as well as classifiers. We introduce a collection of experimental studies derived from the general objectives and motivations of this thesis. The chapter continues with a description of the typical classifier architecture widely adopted in face recognition, namely nearest neighbor classification rule. The chapter also provides an insight into the adopted face databases. We justify their relevance to our research based upon their particular characteristics. The final part of the chapter concentrates on the description of two image transformations, namely contrast enhancement and edge detection. They are used as pre-processing steps in some parts of our research. Some other particular architectures and image manipulation algorithms were considered in some of our investigations. They are introduced in the chapters describing such studies.

In our work we do mimic as much as possible the problems of face recognition outlined in section **Error! Reference source not found.**2.4. In what follows we expand on the details of the experimental scheme and experimental work.

3.1 Experimental scheme

The experimental scheme is designed to provide common ground for our investigations. We devise a research methodology that allows us to evaluate and compare various methods of face recognition. We intend to reveal the particular characteristics (strengths and weaknesses) of some methods capable of constructing meaningful feature spaces. We expose the scenarios in which the face classifiers are more likely to thrive or fail. Our approaches to face recognition fall under the category of individual identification. Therefore the systems under investigation attempt to single out the identity of particular individuals from a group of faces. The assessment of the architectures is expressed in terms of error rates and standard deviations. The error rates express the percentage of misclassified images of a given test set.

The experimental environment contemplates two major datasets regarded as benchmarks, such as FERET and YALE. These datasets are widely used by other researchers in the field, allowing us to compare our findings with those of other laboratories. Details about the datasets are presented later in this chapter. Both face databases portray useful characteristics for our investigations. They include images with variations in lighting and gestures. The FERET database includes a large number of individuals (200) and a much smaller number of images per person (only 3). The FERET database intends to reproduce the small sample size problem - a realistic scenario in most face recognition systems. The YALE database provides numerous images per person (11), and less number of individuals (15). The intrinsic consequence is a better representation of the individuals in the feature spaces. Hence FERET and YALE allow us to evaluate the face classifiers at both extremities of the spectrum.

The experimentation is carried out taking into consideration all the individuals of a database. The recognition systems do not discriminate between known and unknown individuals, but rather they always propose a class for a particular image. Therefore all classes are present in both, training and testing phases.

Three independent datasets with no overlap are built using the images of FERET, and each dataset contains all individuals in it. There are 10 random splits created using images of the YALE database. Once again they include all individuals. In the case of YALE, the random selection of images does allow overlap. The images from FERET and YALE are selected following a uniform random distribution. Hence they all contain approximately the same number of images with facial expressions, lighting conditions, and individuals wearing glasses (in the datasets where such a case applies).

Some experiments documented in this thesis require only training and testing sets. However, there are some cases where the research requires training, validation, and testing sets. Table 1 shows the image distributions for experiments that consider either FERET or YALE. The various sets are kept the same throughout the entire research process, so the comparison between the different architectures is unbiased.

Table 1 Image distribution of training, validation, and testing sets

	Training set	**Validation set**	**Testing set**
FERET	400 images (2 of each person with no overlap)		200 images not included in the training set (1 of each person)
	200 images (1 of each person)	200 images (1 of each person)	200 images (1 of each person)
YALE	90 images (5 of each person)		75 images (4 of each person)
	60 images (4 of each person)	60 images (4 of each person)	45 images (3 of each person)

The architectures that employ images of the FERET database are evaluated by a 3-fold cross validation. In other words, the experiments are repeated 3 times using the 3 combinations of the training-testing or training-validation-testing sets. The combinations of images are never repeated in other sets. The final results are presented as the average performance over the 3 repetitions. The assessments of the experiments utilizing images of YALE are expressed as an average performance over the 10 splits. The experiments consider 10 random splits with overlap, and the final average performance is computed over the 10 repetitions.

The objectives of this research, as outlined in section Chapter 11.2, are: to investigate current and new approaches to face recognition, to identify strengths and weaknesses of different methodologies, and to propose design guidelines and new systems that cope with current limitations. The clearly defined objectives are:

- *Assessing linear and non-linear methods for dimensionality reduction.*
- *Quantifying the impact of image quality (visual deterioration) on the performance of face recognition.*
- *Modeling the effect of image transformation (e.g. edge detection) in classification.*
- *Measuring the influence of image resolution in the performance of classifiers.*
- *Expanding on aggregation of classifiers for combining distinct features.*
- *Investigating on evolutionary optimization in face recognition.*

To accomplish our objectives, we developed a comprehensive suite of experiments that consider various face classifiers. We include a variety of linear and non-linear methods for the construction of meaningful feature spaces. The methods come from the field of statistics and pattern recognition. We also investigate evolutionary methods such as Genetic Algorithms for improving classification. Specific methods under investigation are:

- *Eigenfaces*: it is a linear method based on principal component analysis. It takes advantage of the variances among pixels.
- *Fisherfaces*: it is a linear method based on linear discriminant analysis. It takes advantage of the differences between individuals in order to separate the classes.
- *Kernel-PCA*: it is a non-linear method. Its main property is the capability of calculating non-linear variances of the data.
- *Isomap*: it is a non-linear method. Its goal is to unfold the data onto a lower number of dimensions. It finds the intrinsic geometry (structure) of the data in a given feature space. It accomplishes its task by means of geodesic distances.
- *Genetic Algorithms*: it is an optimization tool. We present a method capable of revealing the importance of each variable from the classification point of view. Our method also finds a suitable similarity measure to distinguish individuals in a given feature space. Evolution is driven towards reducing error rates.

The formal descriptions of the methods outlined above are explained in detail in Chapter 4 and in Chapter 9.

The performance of the classifiers depends directly on the number of variables they take into consideration. In order to find an adequate number of variables, we increased the number of features taken into account by each classifier. The number that led to the lowest error rate was then selected as proper for that given classifier. In this study, the reported error rates are the lowest we computed considering a particular space and number of variables.

Bearing in mind our objectives and the variety of methods, we devise and enumerate several lines of research as follows:

1. *Evaluation of Image distortions in face classifiers*. We evaluate the performance of two widely implemented face classifiers, namely Eigenfaces and Fisherfaces, in the presence of deterioration of visual information. The forms of noise are chosen to mimic real world situations, such as Gaussian noise, salt and pepper, and blurring effect.

Various intensities of such disturbances are taken into consideration. In this line of study we reveal and quantify the relationship that exists between visual disturbances and the performance of the classifiers. Several distance models derived from the Minkowski family of distances are also investigated with respect to the produced classification rates. The findings of our study are important to identify at which levels of noise the face classifiers can still be considered valid. We offer design guidelines and useful recommendations towards improving recognition performance. Details on these evaluations are presented in Chapter 5.

2. *Investigation of the modular approach to face recognition.* We investigate a technique that intends to overcome the limitations of face recognition due to variations of lighting and facial expressions. We describe a method based on Eigenfaces that follows a modular approach for face recognition. We refer to it as modular Eigenfaces or modular PCA. The modular attribute of this approach intends to emphasize particular facial features important for classification. We compare the behavior of modular PCA with that of Eigenfaces. We offer comments on their qualities and weaknesses, and offer guidelines for system design. This investigation is described in Chapter 6.
3. *Impact of image quality (resolution) in face classifiers.* We report on the effect of image resolution and image transformations in face recognition. Image transformations include contrast enhancement via histogram equalization, and edge detection by means of the Sobel operator. Various methods for constructing feature spaces are evaluated, namely Eigenfaces, Fisherfaces, kernel-PCA, and Isomap. Through extensive experimentation, we reveal and quantify the tradeoff between image resolution and transformation, against the performance of the classifiers. The fact that this study takes into consideration several methods capable of dimensionality reduction, image transformations, and image resolutions make this contribution unique. The findings of this research portray practical implications to system design. The complete investigation is reported in Chapter 7.
4. *Aggregation of classifiers based on image transformations.* We present a certain aggregation of features emerging from regular images, edge images, and histogram-equalized images. The features are computed using several methods that prove useful for constructing feature spaces, such as Eigenfaces, Fisherfaces, Isomap, and Kernel-PCA. The findings of our research on the impact of image quality in face classifiers helped us decide on an appropriate image resolution without loosing accuracy, and yet reducing computational cost. To our knowledge, this is the first investigation putting together a comprehensive assessment of numerous face classifiers and combined experts in the presence of image transformations. This is, considering both, linear and non-linear methods for the construction of feature spaces. We present evidence of classification improvement over conventional methods. We also offer design guidelines and recommendations for constructing systems that exhibit robustness against lighting disturbances. Our experimentations and findings are described in Chapter 8.
5. *Evolutionary optimization in face classification.* This is an attempt to improve classification rates in face recognition. Face classifiers commonly select features by ranking them according to their particular variances. However, the variances may not reflect the true importance of such variables. We present a method capable of revealing the importance of each variable from the classification point of view. Our method also finds a suitable similarity measure to distinguish individuals in the given feature space.

Evolutionary optimization is carried out by Genetic Algorithms. Experimental evidence supports the usefulness of this approach in various scenarios. We also comment on major advantages and potential limitations of the proposed architecture. Complete details of this investigation are found in Chapter 9.

3.2 Classification in face recognition – A nearest neighbor approach

The classification of faces is performed in the typical manner found in the area of face recognition. A test image is labeled as belonging to the class depicted by the closest neighbor (face) in a given feature space. We introduce the following notation as a formal description of the classifier.

3.2.1 K-nearest neighbor classifier

In the k-Nearest Neighbor (k-NN) method, a number of patterns K is fixed within a region, therefore the volume of the region varies depending on the data [16]. Given a training set of N observations X={$\mathbf{x}_1$, .., $\mathbf{x}_N$} labeled by c classes and containing N_i patterns in each class c_i, where i = 1, 2, …,C; and $\sum_{i=1}^{c} N_i = N$. A new pattern $\mathbf{x}$ is assigned to a class c_i most frequently appearing within the k-nearest neighbors of $\mathbf{x}$. In other words, for a given $\mathbf{x}$ the first k-nearest neighbors from a training set should be found regardless of the class label, and based on a defined pattern distance measure. Among the selected K nearest neighbors, a majority voting process is performed to assign the appropriate class c_j [16].

As discussed in [84], the k-NN rule can be critically dependent upon the distance used in the construct, especially if there are few examples or K assumes high values. Experimentation concerning different distance functions is required to find the most adequate data for the problem at hand.

3.2.2 Nearest neighbor classification rule

The classification task is cast in the standard framework of pattern recognition. Given the specificity of the task at hand, it is instructive to consider one of the simplest forms of classifiers such as a Nearest Neighbor classification rule (NN classifier). We may envision that the nearest neighbor classifier could serve as the reference classifier. In a nutshell, while the NN- classifier is architecturally very simple and intuitively appealing, a distance function used therein plays an important role. In general, we would be concerned with the exploration of various distance measures. The Minkowski distance takes on the form [16]:

$$d\left(\mathbf{x}',\mathbf{x}\right)=\sqrt[p]{\sum_{j=1}^{n}\left|x_j'-x_j\right|^p} \tag{1}$$

where x_j is the j-th component of the feature vector $\mathbf{x}$. In particular, if p = 1 the relationship arises as the Hamming distance, if p = 2 (1) returns the Euclidean distance. Finally, if p = ∞ we arrive at the Tschebyshev distance.

For the nearest neighbor classification rule, let us consider for instance a set of N fisher feature pairs $(v_1, c_1), \ldots, (v_N, c_N)$, where v_i' is a feature vector in some feature space and c_i takes values in the set $\{1, 2, \ldots, M\}$. Each c_i is considered to be an index of the class which the i-th pattern belongs to, We call $\mathbf{v}_n^{'} \in \{\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_N\}$ a nearest neighbor of v if the following relationship holds: $\min_i\ d(\mathrm{v}_i, \mathrm{v}) = d(\mathrm{v}_n^{'}, \mathrm{v})$ for $i = 1, 2, \ldots, n$, where d denotes a distance model. The nearest neighbor rule assigns v to belong to category $c_n^{'}$ of its nearest neighbor $\mathbf{v}_n^{'}$ [18].

3.3 Face databases

Our experimentation takes into account two well-known face databases regarded as benchmarks in the area of face recognition, namely FERET and YALE. Both datasets include variations in lighting conditions and facial expressions. In what follows, we elaborate on the main features of these two databases.

3.3.1 The FERET database

The faces come from the Face Recognition Technology (FERET) program database of facial images. The FERET evaluation procedure is an independently administered test of face-recognition algorithms. The test serves several purposes: 1.- allows a direct comparison between different algorithms, 2.- identifies the most promising approaches, 3.- assesses the state of the art in face recognition, 4.- identifies future directions of research, and 5.- advances the state of the art in face recognition [81]. The facial images were collected in 11 sessions from August 1993 to December 1994. Conducted at George Mason University and at US Army Research Laboratory facilities, the session lasted one or two days, and the location and setup did not change during a session [81]. The FERET dataset has been widely used and referred to in the face recognition area, see [1][26][60][79][110][111] for examples.

In our work we consider three images per person from a group of 200 individuals, making a total of 600 8-bit grayscale images of 256x384 pixels. In this study, the images were down-sampled by a factor of 2 and trimmed to show only the face area, as presented in Figure 3. The down sampled and trimmed images are 80x100 pixels by 8 bits per pixel. The complete gallery is depicted in Appendix A.

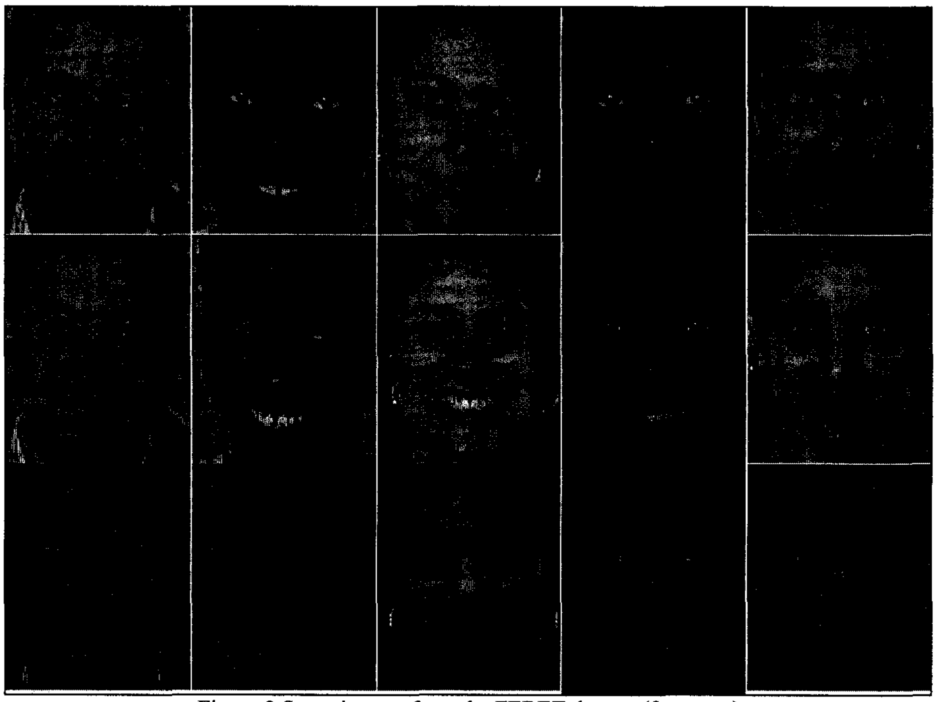


Figure 3 Some images from the FERET dataset (face area)

3.3.2 The YALE database

The Yale Face Database contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. The database is publicly available for non-commercial use in [120]. Other researchers in the field have extensively used the YALE database. Refer to [10][14][58][65] for examples. For the purposes of our experimental work, the images were trimmed to show the face area only, the final image size is 144 x 150 pixels. Some samples of the images are depicted in Figure 4.

Figure 4 Some images from the Yale dataset (face area)

Both face databases used in this study portray useful characteristics for our investigations. The FERET database includes a large number of individuals (200) and a small number of images per person (only 3); such an arrangement intends to reproduce the small sample size problem - a realistic scenario in most face recognition systems. The YALE database provides numerous images per person, with the intrinsic consequence of a better representation of individuals in the feature spaces. The complete collection of images is presented in Appendix A.

3.4 Image transformations

Image transformations constitute a powerful set of operators in digital image processing. In face recognition they are widely used to balance out lighting, remove noise, and stress out different shapes. For the purpose of face recognition, we focus on contrast enhancement by histogram equalization and on edge detection by the Sobel operator. We envision that contrast enhancement can deal, up to some degree, with variations due to illumination conditions, variations that exhibit high variances in many feature spaces. Edge detection is also known to overcome considerable problems of lighting conditions, especially when it comes to recognizing shapes. In the case of face recognition, it can provide more distinctive information regarding the outlines of the face and facial features of each individual, leading to high variances between persons during the process of constructing a feature space.

3.4.1 Histogram equalization

In the histogram equalization process, the gray level values of an image are spread out over the entire gray level plate; hence equal numbers of pixels are allocated to each gray level value. For human observers, this yields more balanced and better-contrasted images. Furthermore, equalized images make details visible in dark or bright regions of the original images [17].

Each pixel of an image **z** assumes an integer value between 0 and 255 for the case of 8-bit gray level plates. Let $\mathbf{h} = [H_0, H_1, ..., H_L]^T$ be the histogram of the enhanced image, where H_l is the number of pixels having value l. In histogram equalization we require that all elements in **h** assume equal values. The pixels in **z** are then reordered into L groups, such that j has H_j pixels. Then all pixels in group j are reassigned gray level value j. Figure 5 illustrates images of the FERET and YALE databases transformed by histogram equalization. The same images without any manipulation are depicted in Figure 3 and Figure 4.

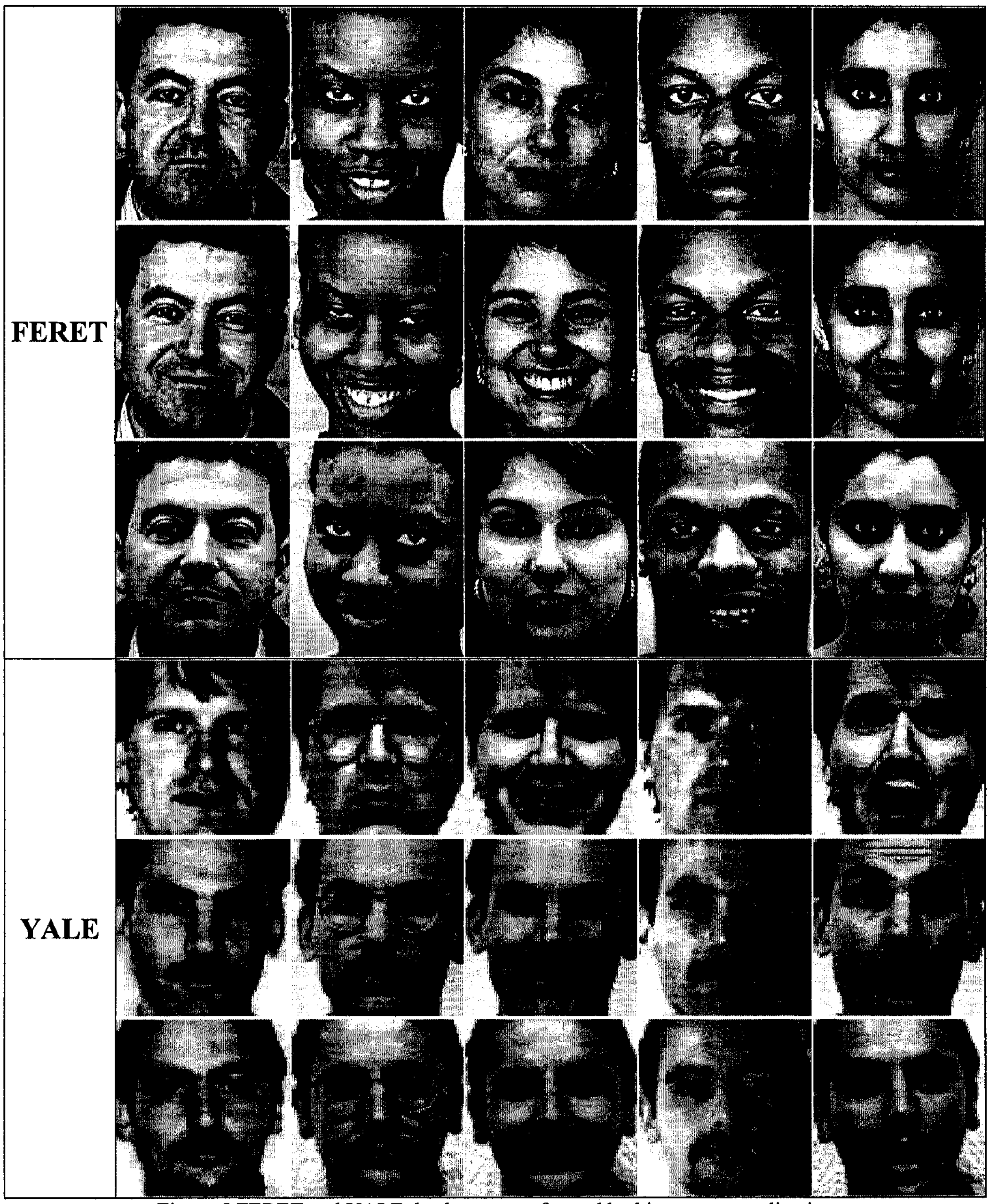


Figure 5 FERET and YALE databases transformed by histogram equalization

Figure 6 depicts some histograms of images with and without contrast enhancement. The illustrations show an even distribution of pixel values along the entire gray level plate for contrast-enhanced images.

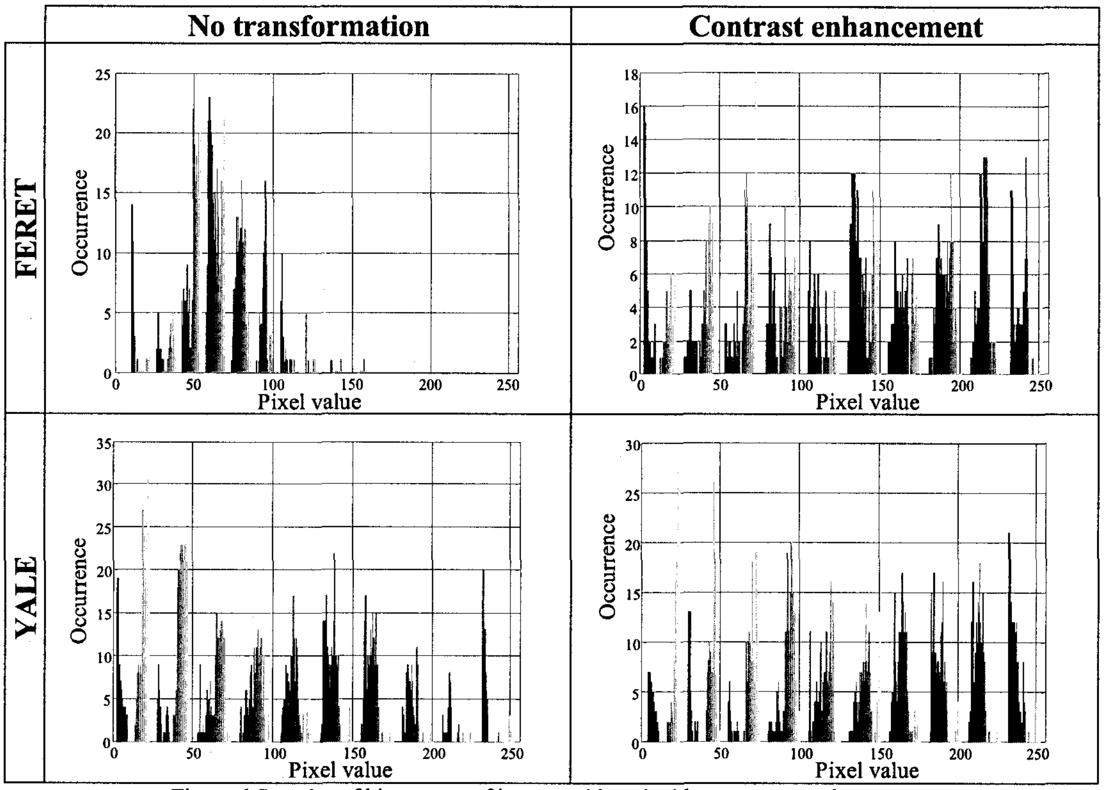


Figure 6 Samples of histograms of images with and without contrast enhancement

3.4.2 Edge detection

The Sobel operator approximates a spatial gradient of an image. Typically, it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image **z**. The Sobel edge algorithm uses a couple of 3x3 convolution masks (Sobel operators), one estimating the gradient in the x-direction and the other estimating the gradient in the y-direction. A mask slides over the image, calculating the gradient of a square of pixels at a time. The typical 3x3 Sobel masks are [20]

-1	0	+1
-2	0	+2
-1	0	+1

Ex

+1	+2	+1
0	0	0
-1	-2	+1

Ey

The gradient magnitude is computed as [42]:

$$Mag = \left(Ex^2 + Ey^2\right)^{1/2} \qquad (2)$$

The resulting image portrays the edges more evidently than in the original image. If a binary image is desired, then a threshold can be applied to the edge image. However, the threshold value may have to be obtained experimentally to achieve satisfactory results. Contrast enhancement can be used prior to edge detection to emphasize the edges in an image. Figure 7 depicts samples of FERET and YALE databases following this technique. The same images without any manipulation are displayed in Figure 3 and Figure 4.

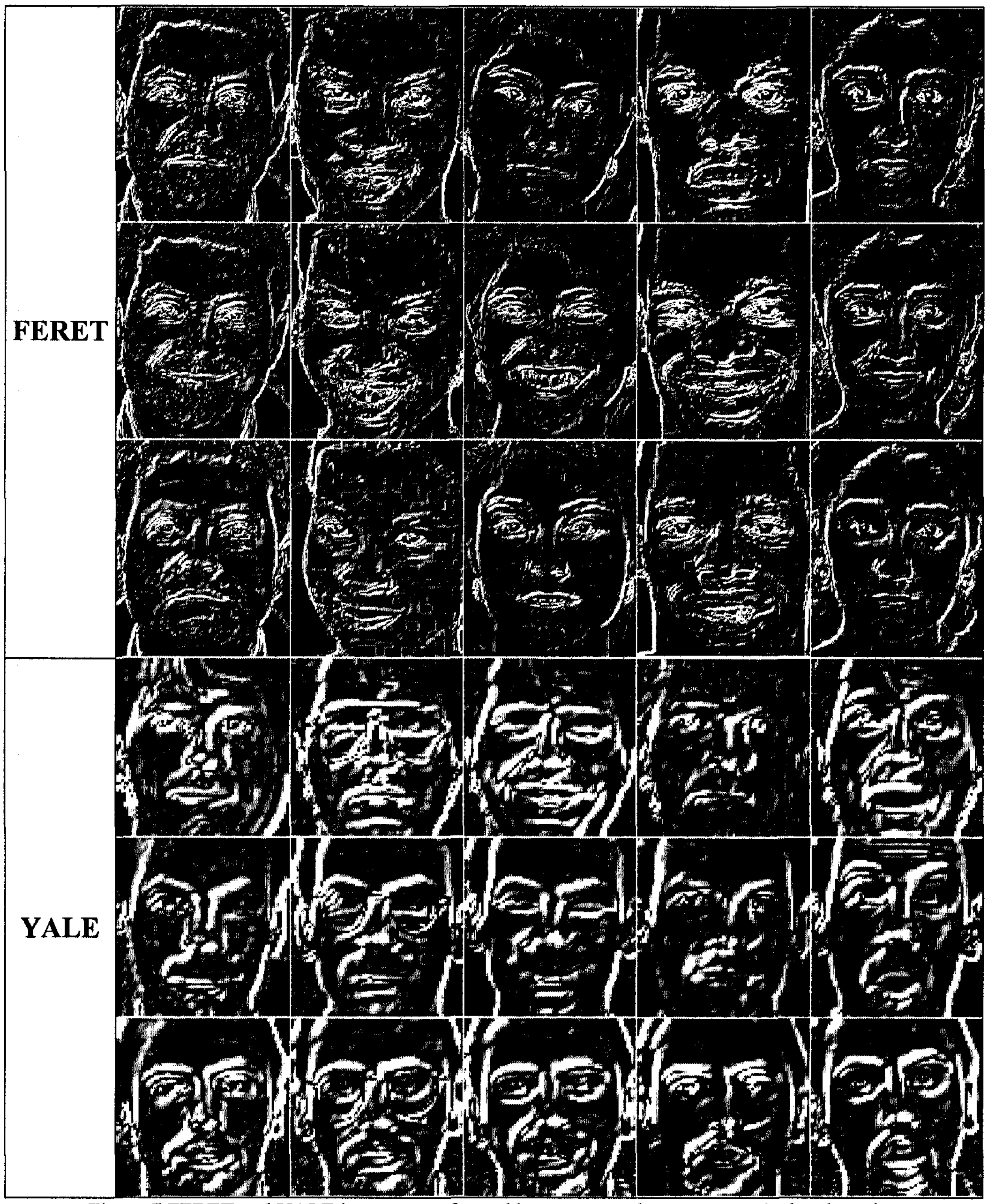


Figure 7 FERET and YALE images transformed by contrast enhancement and edge detection

3.5 Conclusions

This chapter has described the fundamental challenges of face recognition. It has delivered a detailed description of the research methodology and the research schemes adopted in our investigations. A formal description of the classification architecture, namely the nearest neighbor rule has been described. We devised various lines of research that aim at fulfilling the specified objectives. The motivations and contributions of each experimental endeavor have been outlined and justified. The experimental environment has been defined in terms of the face databases and image transformations taken into consideration in this research.

Chapter 4

FEATURE SPACES IN FACE RECOGNITION

This chapter introduces the required theoretical background on the construction of some meaningful feature spaces and image representations in face recognition. The chapter begins by providing a glance at the underlying concept behind dimensionality reduction of input spaces. The chapter continues with the description of some linear and non-linear methods capable of sound dimensionality reduction, such methods have been adopted for our investigations. The methods under consideration are: Eigenfaces, Fisherfaces, Kernel-PCA, and Isomap. We comment on their capabilities and limitations within the context of face representation and reconstruction.

4.1 The making of meaningful features – The underlying concept

One fundamental step in data processing is to reduce the number of variables in a given pattern by extracting only its most informative features. Reduction of pattern dimensionality may improve the recognition process by considering only the most important data. The selected data may include uncorrelated variables that retain most of the information from the original data and that portray best generalization abilities [16]. Dimensionality reduction of the original samples generally refers to a transformation of n-dimensional patterns into some other m-dimensional patterns, where $m \leq n$. The transformation is achieved by a non-linear mapping of the form

$$\mathbf{x}=\mathbf{F}(\mathbf{z}) \qquad (3)$$

where $\mathbf{z}$ denotes the original pattern and $\mathbf{x}$ denotes the new transformed feature. For simplicity, the structure of $\mathbf{F}$ is sometimes chosen to be linear, nonetheless it may still achieve sound dimensionality reduction.

In what follows we describe some linear and non-linear models for dimensionality reduction (construction of feature spaces). They have proven useful in many areas including biometrics. We refer to Eigenfaces, Fisherfaces, kernel-PCA, and Isomap.

4.2 Eigenfaces

Principal Component Analysis (PCA) is perhaps the most popular method for reducing the number of variables in face recognition. In PCA, frequently called "Eigenfaces" in face recognition, faces are represented as a linear combination of weighted eigenvectors [107]; the basis functions are the eigenvectors of the covariance matrix of a training image set [12]. Eigenfaces takes advantage of the similarity between the pixels among images by means of their covariance matrix. In this regard we argue that a linear relationship (correlation) exists among neighboring pixels, for example, the pixels representing the areas of the forehead may present similarities among themselves. On the other hand, pixels related to the mouth or eyes may contain relevant information that is unique to an admitted extent.

To accomplish the training process of face classifiers in the framework of Eigenfaces, it becomes necessary to compute eigenvectors and eigenvalues of the covariance matrix of the training image set. These eigenvectors define a new face space where the images are represented. Given the computed eigenvectors, we construct a set of feature vectors for each image. To fix the required notation, let us introduce the following symbols.

Let an image, coming in the form of some array of x by y pixels, be represented as a single vector $\mathbf{z}$ of n inputs. Given a set Z = {$\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$} of N image vectors formed in an n dimensional space, where each element of Z belongs to a certain class, the covariance matrix R is defined in a usual manner

$$\mathrm{R} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T \tag{4}$$

where $\bar{\mathbf{z}}$ is the mean vector of all images in the training set. Let us denote the term $(\mathbf{z}_i - \bar{\mathbf{z}})$ in (4) by Ψ_i, and $\mathrm{A} = [\Psi_1,\ldots,\Psi_N]$. Then R can be expressed as

$$\mathrm{R} = \frac{1}{N}\mathrm{A}\cdot\mathrm{A}^T \tag{5}$$

Since $\mathrm{R} \in \mathbf{R}^{n\times n}$, the number of entries of R becomes an issue. The computation becomes infeasible and expensive even for small images. The reduction of the number of variables is a must. Fortunately we can achieve reduction by considering a smaller matrix $\mathrm{R} \in \mathbf{R}^{N\times N}$ [98] instead. Given $\mathrm{B} = \mathrm{A} / \sqrt{N}$, we form a new matrix T such that

$$\mathrm{T} = \mathrm{B}^T\mathrm{B} \tag{6}$$

where $\mathrm{T} \in \mathbf{R}^N$, Notably, the eigenvalues of R are the same as the eigenvalues of T, and the eigenvectors of R are the same as the normalized eigenvectors of TB. B can be decomposed into $\mathrm{B}=\mathrm{UWV}^T$ by computing the Singular Value Decomposition (SVD). By substitution of B in (6), T becomes

$$\mathrm{T} = \mathrm{VW}^T\mathrm{WV}^{\mathrm{T}} \tag{7}$$

where V is the matrix of eigenvectors obtained from T. By definition of eigenvectors, if $\mathbf{v}_i$ is an eigenvector corresponding to an eigenvalue λ_i, we can develop $\mathrm{T}\mathbf{v}_i = \lambda_i\mathbf{v}_i$, therefore $\mathrm{B}^T\mathrm{B}\mathbf{v}_i = \lambda_i\mathbf{v}_i\mathrm{T}\mathbf{v}_i = \lambda_i\mathbf{v}_i$. After some manipulation we arrive at $\mathrm{RB}\mathbf{v}_i = \lambda_i\mathrm{B}\mathbf{v}_i$. $\mathrm{B}\mathbf{v}_i$ represents the i-th eigenvector of the covariance matrix R. Therefore, given that V is the matrix of eigenvectors of R, it follows that BV=E. The eigenvalues of R are computed as W^TW. This new formulation allows us to solve the eigenvalue problem

$$\mathrm{TE} = \lambda \mathrm{E} \tag{8}$$

where λ denotes the eigenvalues of T, and E stands for the corresponding eigenvectors. Thus, the presented approach is feasible and far less expensive to compute the eigenvalues and eigenvectors of the covariance matrix R. Note that the maximum number of eigenvectors that we can obtain in this way is at most N. If the eigenvectors are depicted as images they portray a set of ghostly faces, commonly known as Eigenfaces.

Once the face space has been constructed, the feature vectors are formed as a linear combination of the eigenvectors of the covariance matrix R. We project an image $\mathbf{z}_i$ into the face space through the following transformation

$$\mathbf{x}_i = \mathrm{E}^T(\mathbf{z}_i - \overline{\mathbf{z}}) \tag{9}$$

where $\mathbf{x}_i$, i=1,...,N are the weight vectors associated with the eigenvectors in E. One can experiment with the number of eigenvectors to compute the weights, generally only a few amount provide sufficient information for adequately representing the images in the face space and for reconstruction purposes.

4.3 Fisherfaces

Linear Discriminant Analysis (LDA) comes as an extension of the Eigenfaces method in the sense that it is more robust against lighting conditions and facial expressions [39]. LDA, i.e. "Fisherfaces" in face recognition, is a class-specific approach in the sense that it attempts to maximize the separability of the classes within the linear subspace. Fisherfaces maximizes the ratio of the scatter between classes and the scatter within classes in order to form a new projection space. Fisherfaces takes advantage of the fact that, under ideal conditions, the variation within class falls in a linear subspace of the image space [10]. Hence, the classes are convex, and therefore, linearly separable [10]. One can complete the reduction of variables by using linear projection and still preserve linear separability. This is a strong argument in favour of using linear methods for dimensionality reduction, at least when seeking robustness against lighting conditions [10].

To find the scatter within classes and the scatter between classes it is necessary to reduce the number of variables of the original images in Z beforehand. To this end one can follow PCA. Let $\mathbf{x}$ be a projected image positioned in the face space. The scatter matrix between classes reads as

$$\mathrm{S_B} = \sum_{j=1}^{c} N_j(\overline{\mathbf{x}}_j - \overline{\mathbf{m}})(\overline{\mathbf{x}}_j - \overline{\mathbf{m}})^T \tag{10}$$

where C is the number of classes, N_j is the number of images belonging to class j, $\overline{\mathbf{x}}_j$ is the mean vector of imaged transformed by PCA belonging to class j, and $\overline{\mathbf{m}}$ is the mean vector of all images transformed by PCA. The scatter matrix within classes is defined as

$$\mathrm{S_W} = \sum_{j=1}^{c} \sum_{\mathbf{x}_i \in j} (\mathbf{x}_i - \overline{\mathbf{x}}_j)(\mathbf{x}_i - \overline{\mathbf{x}}_j)^T \tag{10}$$

The ratio to be maximized comes in the form

$$\mathrm{D_{LDA}} = \arg\max_{\mathrm{D}} \frac{\left|\mathrm{D}^T \mathrm{S_B} \mathrm{D}\right|}{\left|\mathrm{D}^T \mathrm{S_W} \mathrm{D}\right|} = \left[\mathbf{d}_1, \ldots, \mathbf{d}_{c-1}\right]^T \tag{11}$$

where $\mathrm{D_{LDA}} = [\mathbf{d}_1,..,\mathbf{d}_{c-1}]$ denotes the generalized eigenvectors of $\mathrm{S_B}$ and $\mathrm{S_W}$ associated to the largest eigenvalues λ_q, q = 1,..,C-1, In other words $\mathrm{S_B}\mathbf{d}_q = \lambda_i \mathrm{S_W} \mathbf{d}_q$, q = 1,...,C-1, meaning that

$$\mathrm{S_W}^{-1} \mathrm{S_B} \mathrm{D_{LDA}} = \lambda \mathrm{D_{LDA}} \tag{12}$$

Note that the rank of $\mathrm{S_B}$ is C-1 since we have at most C-1 summations in the scatter between class matrix. To construct a Fisher feature vector $\mathbf{v}_i$ we form a new linear combination of the features transformed by PCA with $\mathrm{D_{LDA}}$, this is

$$\mathbf{v}_i = \mathrm{D_{LDA}}^T \mathbf{x}_i = \mathrm{D_{LDA}}^T \mathrm{E}^T (\mathbf{z}_i - \overline{\mathbf{z}}), \text{ for } i = 1,..,N. \tag{13}$$

4.4 Kernel-PCA

Unlike PCA, kernel-based methods exploit non-linear relationships among patterns by performing a non-linear form of principal component analysis. The use of integral operator kernel functions allows computation of principal components in highly dimensional feature spaces instead of in the original space. The selected non-linear kernel model relates the highly dimensional space to the input space. Schölkopf, *et al.* [90] extended principal component analysis to a non-linear form based on kernel methods. Given the centered images $\mathbf{z}_i$, i=1,...,N, $\mathbf{z}_i \in \mathbf{R}^n$, $\sum_{i=1}^{N} \mathbf{z}_i = 0$, we compute PCA in another inner-product space F, which is related to the input space by a possibly non-linear mapping Φ,

$$\Phi : \mathbf{R}^n \rightarrow F, \mathbf{z} \mapsto \mathbf{Z} \tag{14}$$

Where F denotes the higher dimensional space, which could be very large, possibly infinite. Again, assuming centered data, that is $\sum_{i=1}^{N} \Phi(\mathbf{z}_i) = 0$, the covariance matrix in F is

$$R = \frac{1}{N}\sum_{i=1}^{N}\Phi(\mathbf{z}_i)\Phi(\mathbf{z}_i)^T \tag{15}$$

We now solve the eigenvalue problem RE = λE. All solutions E with corresponding non-zero eigenvalues lie in the span of $\Phi(\mathbf{z}_1),\ldots,\Phi(\mathbf{z}_N)$. This has two useful consequences: we may instead consider the set of equations

$$\lambda(\Phi(\mathbf{z}_q)\cdot E) = (\Phi(\mathbf{z}_q)\cdot RE)\text{, for } q = 1, \ldots, N. \tag{16}$$

And there are coefficients α_i, i = 1,...,N such that $E = \sum_{i=1}^{N}\alpha_i\Phi(z_i)$, Therefore by substitution in (16)

$$\lambda\sum_{i=1}^{N}\alpha_i(\Phi(\mathbf{z}_q)\cdot\Phi(\mathbf{z}_i)) = \frac{1}{N}\sum_{i=1}^{N}\alpha_i(\Phi(\mathbf{z}_q)\cdot\sum_{j=1}^{N}\Phi(\mathbf{z}_j))(\Phi(\mathbf{z}_j)\cdot\Phi(\mathbf{z}_i))\text{, for } q = 1,\ldots, N. \tag{17}$$

We define an $N \times N$ matrix K, whose elements $K_{st} = (\Phi(\mathbf{z}_s)\cdot\Phi(\mathbf{z}_t))$, hence by substitution in (17) and multiplying both sides by N, (17) can be expressed as $N\lambda K\alpha = K^2\alpha$, or equivalently

$$N\lambda\alpha = K\alpha \tag{18}$$

Let λ denote the non-zero eigenvalues of K, i.e. the solutions $N\lambda$ of (18), and α = $[\alpha_1,\ldots,\alpha_N]$ the corresponding eigenvectors. For the purpose of principal component extraction we need to compute projections onto the eigenvectors E = $[\mathbf{e}_1,\ldots,\mathbf{e}_N]$ in F. Let $\mathbf{z}$ be a test image with a projection $\Phi(\mathbf{z})$ in F; then

$$(\mathbf{e}_q\cdot\Phi(\mathbf{z})) = \sum_{i=1}^{N}\boldsymbol{\alpha}_{qi}(\Phi(\mathbf{z}_i)\cdot\Phi(\mathbf{z}))\text{, for } q = 1,..,N. \tag{19}$$

The previous outlines the general procedure of non-linear PCA, but we have not expanded on some possible non-linear models to do the non-linear mapping Φ. In order to compute dot products of the form $(\Phi(\mathbf{z})\cdot\Phi(\mathbf{y}))$ we use kernel representations, $k(\mathbf{z},\mathbf{y}) = (\Phi(\mathbf{z})\cdot\Phi(\mathbf{y}))$, which allow us to compute the value of the dot product in F without having to carry out the map Φ. Kernel models adopted in our research are:

- *The polynomial kernel*: $k(\mathbf{z},\mathbf{y}) = (\mathbf{z}\cdot\mathbf{y})^d$
- *The Gaussian kernel, also called Radial basis kernel*: $k(\mathbf{z},\mathbf{y}) = \exp\left(-\frac{\|\mathbf{z}-\mathbf{y}\|^2}{2\sigma^2}\right)$
- *The Sigmoid kernel*: $k(\mathbf{z},\mathbf{y}) = \tanh(\kappa(\mathbf{z}\cdot\mathbf{y}) + \Theta)$, where Θ stands for an offset.

4.5 Isomap

Tenenbaum, *et al.* [95] described a method to reduce the number of variables (dimensionality reduction) by using local metric information, and by learning the underlying geometry of the data. The main idea behind Isomap is to overcome the limitations of linearity by calculating non-linear distances between data samples; this is achieved by means of the so-called geodesic or curvilinear distance. The curvilinear distance depends not only on the two points for which the distance is measured, but also on their adjacent points [56]. The objective of the curvilinear distance is to compute distances along an object. For example, a plane cannot fly from New York to Tokyo by following a straight line, instead it has to follow the curvature of the Earth. This simple comparison illustrates why curvilinear distances are also known as geodesic distances [56]. In practical situations only a discrete representation of the data is known, therefore only a discrete approximation of the curvilinear distance can be computed. Instead of measuring the length of the intrinsic curve, this is following the entire geometry formed by all data samples, we sum the lengths of small interconnecting segments that approximate the curve.

The geodesic distance between two points is approximated by finding the neighboring data points of each sample, for instance by means of the k-nearest neighbor algorithm. Once the neighborhoods are known, a graph is built by linking all neighboring points. Each arc of the graph is labeled with the Euclidean distances between corresponding linked points. Ultimately the geodesic distance between two points is approximated by the sum of the arc lengths along the shortest path linking both points [55]. One can compute the shortest path between the two samples by following Dijkstra's algorithm.

Given the features vectors $\mathbf{x}_i$, i=1,..,N, the Isomap algorithms processes the data as follows:

1. *Connect neighboring points*: connect each point either with K closest other ones (following k-nearest neighbors rule), or with those enclosed within a certain radius ε.
2. *Compute a matrix D of all pairwise geodesic distances*: run Dijkstra's algorithm for each point and store the pairwise distances between all points in a symmetric matrix D with $N \times N$ entries.
3. *Center D*: compute the mean of the rows, the mean of the columns and the mean of all entries; subtract the mean of rows from each row, subtract the mean of columns from each column, and the grand mean from all entries.
4. C*ompute the eigenvalues and eigenvectors of the centered matrix D*: sort the eigenvectors according to the descending order of the eigenvalues. The matrix of sorted eigenvalues serves as a non-linear transformation matrix of the original data [56].

An important parameter to specify in Isomap is the number of closest neighbors K, or the radius ε of enclosing neighbors to connect all data samples. If $K = N$ we end up computing regular PCA, if K is too large we may not achieve any reduction in the number of variables, and the data geometry may be too difficult to approximate. The goal is therefore to find the lowest value of K so that all data points are connected.

One major drawback of Isomap is that, in many cases, the data does not easily reveal its geometry. This issue becomes more evident when we deal with large number of variables or when data samples are too close to each other in a given space.

4.6 Conclusions

This chapter has introduced the required notations for the construction of meaningful feature spaces in the context of face recognition. The methods include linear and non-linear methods for the extraction of useful information from the images originally depicted as a collection of pixels. The introduced methods are capable of sound dimensionality reduction in the area of face recognition.

Chapter 5

EFFECTS OF ENVIRONMENTAL DISTURBANCES IN EIGENFACES AND FISHERFACES[1]

There has been an ongoing quest to design systems that exhibit high classification rates and robustness. This feature becomes of paramount relevance when dealing with noisy and uncertain images. The design of face recognition classifiers capable of operating in the presence of deteriorated (noise affected) face images requires a careful quantification of deterioration of the existing approaches vis-à-vis anticipated forms and levels of image distortions.

The objective of this experimental study is to reveal some general relationships characterizing the performance of two commonly used face classifiers (that is Eigenfaces and Fisherfaces) in the presence of deteriorated visual information. The findings obtained in our study are crucial to identify at which levels of noise the face classifiers can still be considered valid. Prior knowledge helps us develop adequate face recognition systems. We investigate several typical models of image distortion such as Gaussian noise, salt-and-pepper, and blurring and demonstrate their impact on the performance of the two main types of the classifiers. Several distance models derived from the Minkowski family of distances are investigated with respect to the produced classification rates.

5.1 Noisy Environment

In this study we are concerned in revealing essential relationships between the deterioration of visual information available to the classifier (images) and the resulting error rates. The forms of noise are chosen to mimic real world situations. We concentrate on three essential mechanisms of information distortion coming at different levels of degradation. We admit the existence of several other important types of noise affecting face recognition, for instance occlusion. At this point we argue that occlusion is not necessarily a type of noise but rather a particular manner in which individuals are presented, in other words, the environment does not play a role in the occlusion of images. In what follows, we briefly characterize the noise models used in this study.

5.1.1 Gaussian noise

Noise having Gaussian-like distribution is very often encountered in acquired data. For example, since an image is literally a flow of energy, it is necessary to use some recording mechanism that captures the energy flow at a particular time. Such mechanisms are always imperfect and add some noise to the images [23]. To simulate this form of noise, the values

[1] A version of this chapter has been published.
Jarillo, Pedrycz, Reformat, Kwak, "Deterioration of Visual Information in Face Classification Using Eigenfaces and Fisherfaces", *Machine Vision and Applications,* Vol. 17, No. 1, pp. 68-82, 2006.

of all pixels are modified to follow a normal Gaussian distribution with a zero mean and certain variance.

5.1.2 Salt and pepper noise

This type of noise is present when an image is coded and transmitted over a noisy channel or degraded by electrical sensor noise [23]. As such it is commonly found in scanned images. Salt and pepper noise consists of white and black dots (salt and pepper) distributed randomly throughout an image. To reproduce this noise, the value of randomly selected pixels is changed to either white or black, and a probability density function specifies the quantity of pixels to be modified. Half of the pixels are changed to white and the remaining half to black.

5.1.3 Blurring

Blurring due to misfocus of lenses, motion, or atmospheric turbulence can degrade an image. Essentially the high frequency details of the image are reduced [23]. In this study, the out of focus blur is considered. The blurring is generated by computing the average of the pixels around a particular pixel. The block size defines the number of side pixels to form the block. The new value is set to the pixel in the center of such block. The blurring effect increases as the block size increases.

To complete a set of experiments, we consider several levels of degradation of visual information by admitting the following scenarios

- Gaussian noise with zero mean and variance of 0.01, 0.02,..., 0.05
- Salt and pepper noise with density equal to 10%, 20%,..., 50%
- Blurring effect with averaging square block of sizes 3x3, 5x5,..., 13x13

Figure 8 shows examples of the FERET dataset with superimposed distortion.

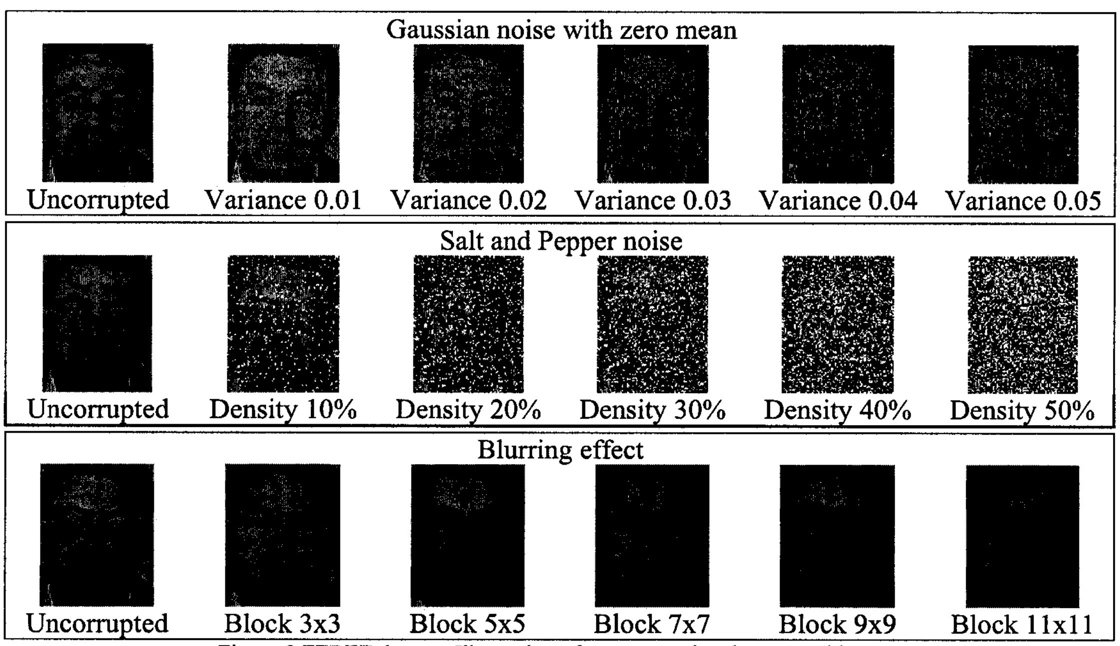


Figure 8 FERET dataset, Illustration of uncorrupted and corrupted images

5.2 Experimental setup

The setup is geared towards a thorough and in-depth investigation of the robustness of the Eigenfaces and Fisherfaces classifiers operating in different noisy environments, involving both, the type of deterioration and its intensity. We report on the performance of the classifier in terms of error rate in the usual manner [86][93][69]. Two different experimental scenarios are considered in our investigation:

- *Scenario 1 - Training with uncorrupted images.* It involves training of Eigenfaces and Fisherfaces classifiers with the use of clean (uncorrupted) images, while the testing is performed with uncorrupted and corrupted images.

- *Scenario 2 - Training with corrupted and uncorrupted images.* For this scenario the training set includes the collection of uncorrupted images mixed with some corrupted images, while the testing is performed over uncorrupted and corrupted images.

The levels of image deterioration occurring in the training sets are referred as low, medium, and high. For details refer to Table 2. To obtain the representative behavior of the classifiers, an average is taken over a number of experiments. The experimental environment considers the FERET database in the standard manner described in section 3.1. A 3-fold cross validation across individuals is performed (two randomly selected images to train and one remaining image to test). All classes are considered in the training and testing sets. The training and testing sets of the scenario 1 comprise 400 and 200 images respectively, while in scenario 2 there are 800 images in the training set (400 uncorrupted images plus the same images corrupted by one type of distortion, and level of deterioration) and 200 images in the testing set. In summary, Table 3 presents the characteristics of both scenarios.

Table 2 Levels of image distortion

Distortion model	**Low distortion**	**Medium distortion**	**High distortion**
Blurring effect	3x3 pixel block	7x7 pixel block	9x9 pixel block
Salt and pepper noise	Density 10%	Density 30%	Density 50%
Gaussian noise	Variance 0.01	Variance 0.03	Variance 0.05

Table 3 Description of training and testing sets

	Training set	Testing set
Scenario 1 - training with uncorrupted images	3 sets of 400 uncorrupted images	3 sets of 200 images. Each level of distortion, including uncorrupted images, is considered independently in the testing sets.
Scenario 2 - training with corrupted and uncorrupted images	3 sets of 400 uncorrupted images combined with 400 corrupted images at low level of distortion	
	3 sets of 400 uncorrupted images combined with 400 corrupted images at medium level of distortion	
	3 sets of 400 uncorrupted images combined with 400 corrupted images at high level of distortion	

As a first step, the Eigenfaces classifier is trained and later tested using only uncorrupted images. The error rates are computed considering different numbers of eigenvectors, starting with the eigenvectors with the highest eigenvalues to the maximum possible number of eigenvectors. Later on, the classifier considers only the selected eigenvectors during the classification process. Eigenfaces is then tested with corrupted images affected by different levels of degradation.

In the same fashion, Fisherfaces is trained and later tested with only uncorrupted images. Subsequently the error rates are computed according to variations in the number of discriminant vectors. This approach allows us to determine the optimal discriminant vectors to compute the feature vectors. Fisherfaces is then used as classifier taking into account only the chosen discriminant vectors. It is later tested with corrupted images and different levels of image distortion, all this following a 3-fold cross validation technique.

The vector size of the images transformed by PCA to train the Fisherfaces method is based upon the portion of the total variance that their eigenvectors account for, in this study Fisherfaces was trained using the portion that account for 90% (or the closest possible) of the total cumulative variance. Other selections of variance have been used to compute the weights of the images, for example, Bartlett, *et al.* [9] considered those eigenvectors that account for 98% of the variance, which is approximately 200 eigenvectors. However, we found that those eigenvectors that account for 90% of the variance provide lower error rates in the case of Fisherfaces method.

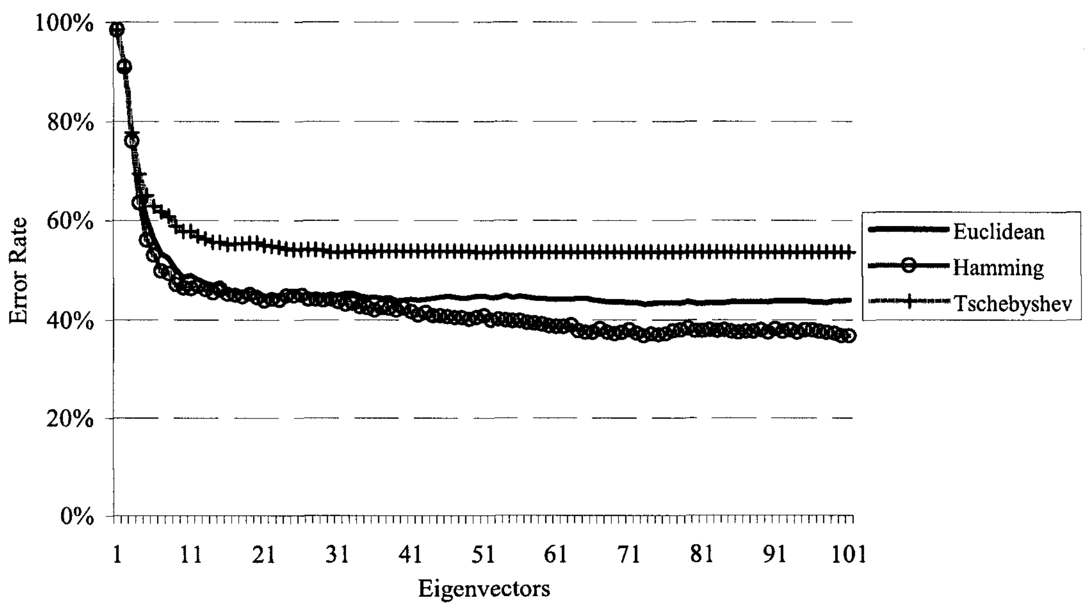
In general, the error rate decreases as we increase the number of eigenvectors and discriminant vectors. It is of interest to find the general adequate number of eigenvectors to represent the images in the face space without damaging the classification rates. This reduces the computational effort of the face classification methods and provides a simpler representation of the images in the face space.

5.3 Experimental results

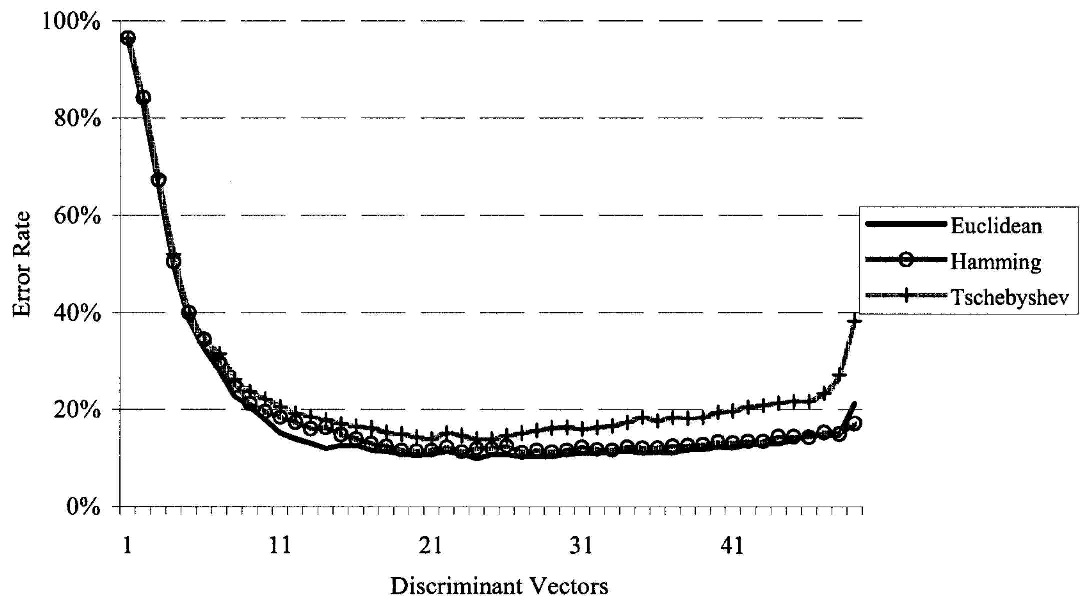
This section reports on the experimental findings of Eigenfaces and Fisherfaces classifiers under the scenarios outlined above. We start with the first scenario (training with uncorrupted images) and then move on to the second one (involving corrupted images in the training phase).

Scenario 1 - Training with uncorrupted images

Figure 9 summarizes the error rates of the Eigenfaces and Fisherfaces methods obtained for the first scenario using several distances. To make the plots readable we show error rates up to 100 eigenvectors. In case of Eigenfaces, the numbers of eigenvectors that provide the lowest error rates are 73, 200, and 30 eigenvectors for the Euclidean, Hamming, and Tschebyshev distance respectively. In the case of Fisherfaces, the classifier was trained using the portion of faces transformed by PCA that accounts for approximately 90% of the total cumulative variance.



a) Eigenfaces



b) Fisherfaces

Figure 9 Average error rates of a) Eigenfaces and b) Fisherfaces classifiers

Please observe that in Fisherfaces there are at most *C*-1 discriminant vectors available, therefore the previous plots depict Fisherfaces up to such extent along the independent axis. For further insight refer to [10]. In Figure 9 we notice that the error rate goes up as the number of discriminant vectors increases, this may suggest that some over-fitting has happened in the training phase. Since there are only two images per class and 200 classes, we fall into the small sample size problem, not having enough images to adequately "cover" the entire Fisherfaces space. Each discriminant vector adds an extra dimension in the fisher space to represent the images. However, most of the variance is comprised in the primary discriminant vectors. As we add more discriminant vectors to compute the feature vectors, we also add extra information that may not be relevant to distinguish between individuals, and at the same time we also expand the dimensionality of the space. If the variances attributed to the added discriminant vectors are small compared to the rest, then we are basically not providing relevant information to the feature vectors. The distance function may offer some insight at this point, for example, the classification using Tschebyshev distance shows the most predominant increase in the error rate. This can be thought of as if the discriminant vectors that produce the elements in the feature vectors are those that do not introduce relevant information for classification, and they could very well be those used to compute the distance, shadowing the relevant ones.

Figure 10 and Figure 11 show the impact of image distortion over the testing sets using Eigenfaces and Fisherfaces classifiers. The data is organized according to types and levels of noise in the testing sets. A positive standard deviation is also depicted on top of each bar. Table 4 refers to the number of eigenvectors and discriminant vectors that provided the lowest error rates in the classification of uncorrupted images, these eigenvectors and discriminant vectors were used to compute the error rates present in the figures.

Table 4 Number of eigenvectors and discriminant vectors delivering low error rates

	Euclidean	Hamming	Tschebyshev
Number of eigenvectors	73	200	30
Number of discriminant vectors	24	27	21

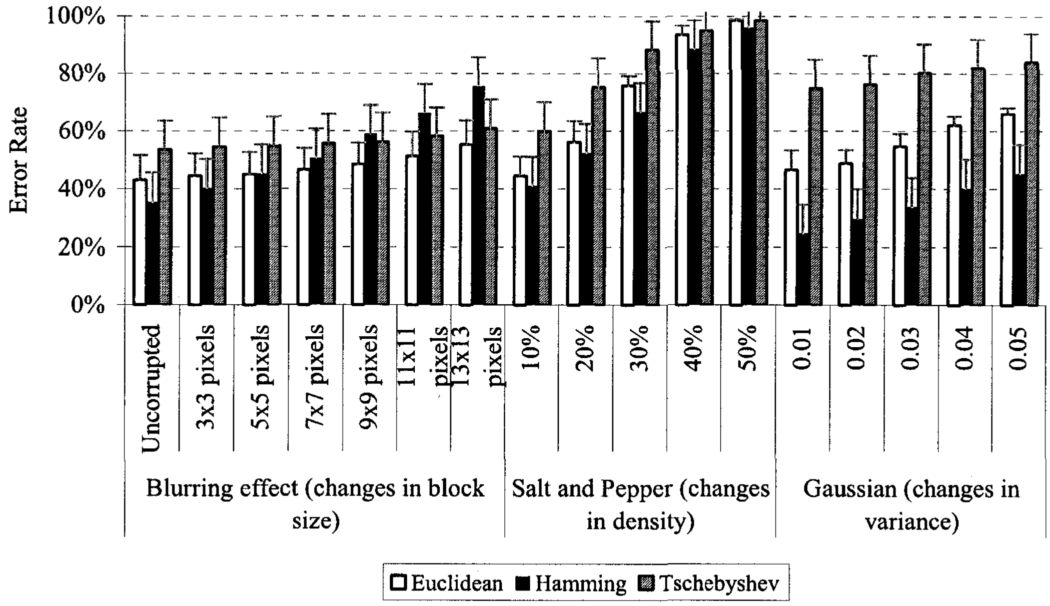


Figure 10 Impact of image distortion on Eigenfaces method

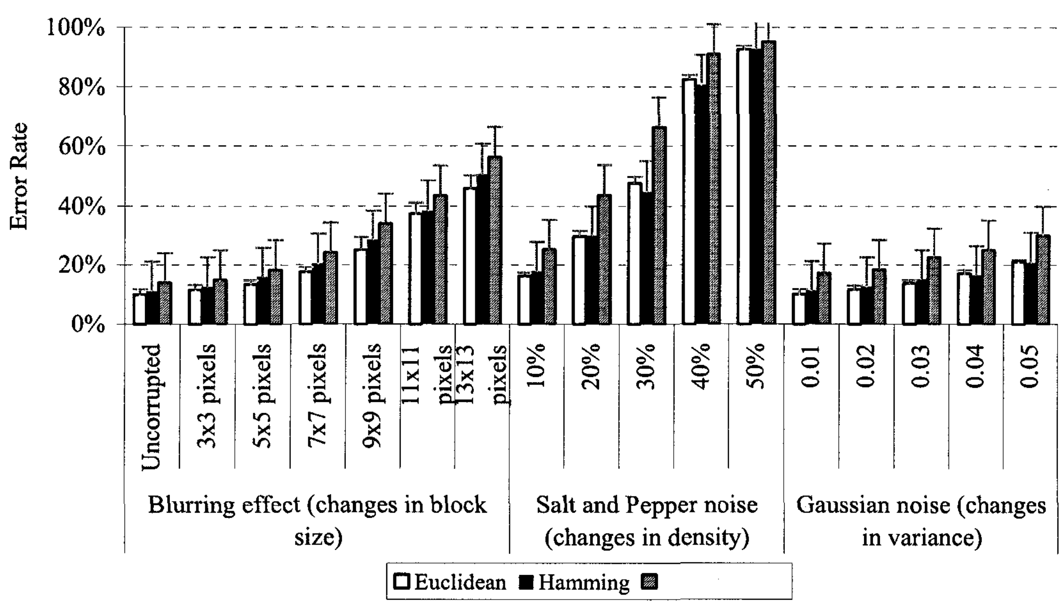


Figure 11 Impact of image distortion on Fisherfaces method

Scenario 2 - Training with corrupted and uncorrupted images

The main results are collected across various image distortion models. Table 5 presents the number of eigenvectors that provide the lowest error rates in Eigenfaces method, the reported results are obtained using a collection of uncorrupted images combined corrupted images with low, medium, and high level of distortion in the training sets.

Table 5 Number of eigenvectors leading to low errors for various deteriorations, Eigenfaces

Distortion model	**Distortion level**	**Distance type**		
		Euclidean	**Hamming**	**Tschebyshev**
Blur	Low	78	358	29
	Medium	61	373	11
	High	145	139	24
Salt and pepper	Low	86	175	31
	Medium	89	247	17
	High	90	206	36
Gaussian	Low	75	155	48
	Medium	71	163	31
	High	71	423	31

Table 6 describes the number of discriminant vectors that provide the lowest error rate over uncorrupted images. These eigenvectors are those to be used by Fisherfaces when carrying out testing with the use of corrupted images.

Table 6 Number of discriminant vectors leading to low errors for various distortions, Fisherfaces

Distortion model	**Distortion level**	**Distance type**		
		Euclidean	**Hamming**	**Tschebyshev**
Blur	Low	18	21	17
	Medium	25	26	11
	High	23	23	12
Salt and pepper	Low	29	29	21
	Medium	10	14	15
	High	6	5	6
Gaussian	Low	26	24	23
	Medium	34	44	31
	High	26	20	28

Figure 12 and Figure 13 show the impact of blurring effect over Eigenfaces and Fisherfaces methods respectively, using Euclidean, Hamming, and Tschebyshev distances. The performance of Eigenfaces and Fisherfaces is scrutinized in terms of error rate and standard deviation. The independent axis shows three main groups labeled as low, medium, and high distortion levels. These are the distortion levels accounted in the training sets; each of these groups presents several distortion levels that correspond to the image distortion in the testing sets. Table 5 and Table 6 denote the number of eigenvectors and discriminant vectors that provided the lowest error rates in the classification over uncorrupted images;

these eigenvectors and discriminant vectors were used to compute the error rates presented in the figures.

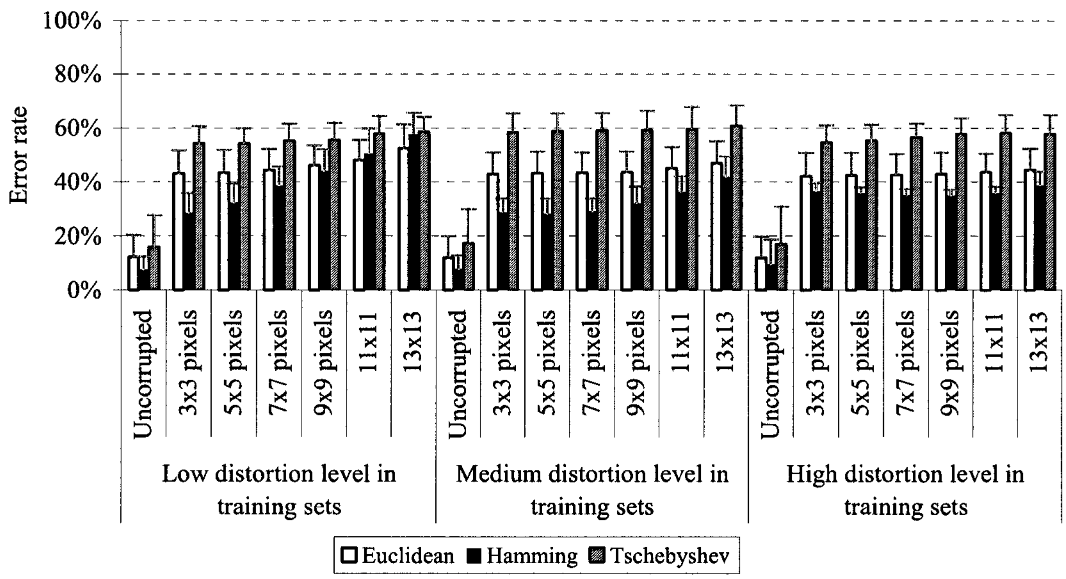


Figure 12 Impact of blurring on Eigenfaces with uncorrupted and corrupted images for training

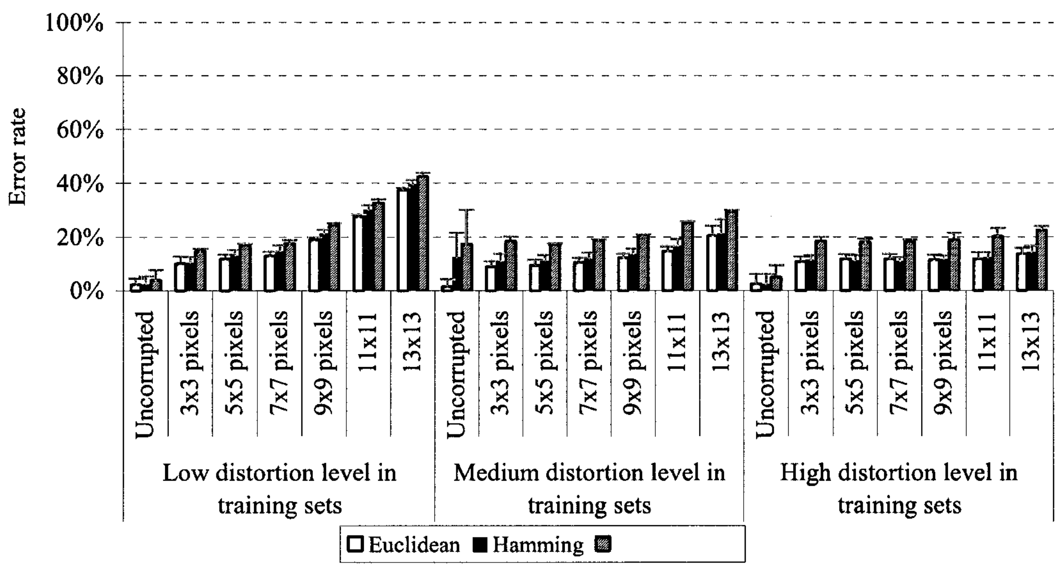


Figure 13 Impact of blurring on Fisherfaces with uncorrupted and corrupted images for training

In the same fashion Figure 14 and Figure 15 present the impact of salt and pepper noise on Eigenfaces and Fisherfaces methods respectively.

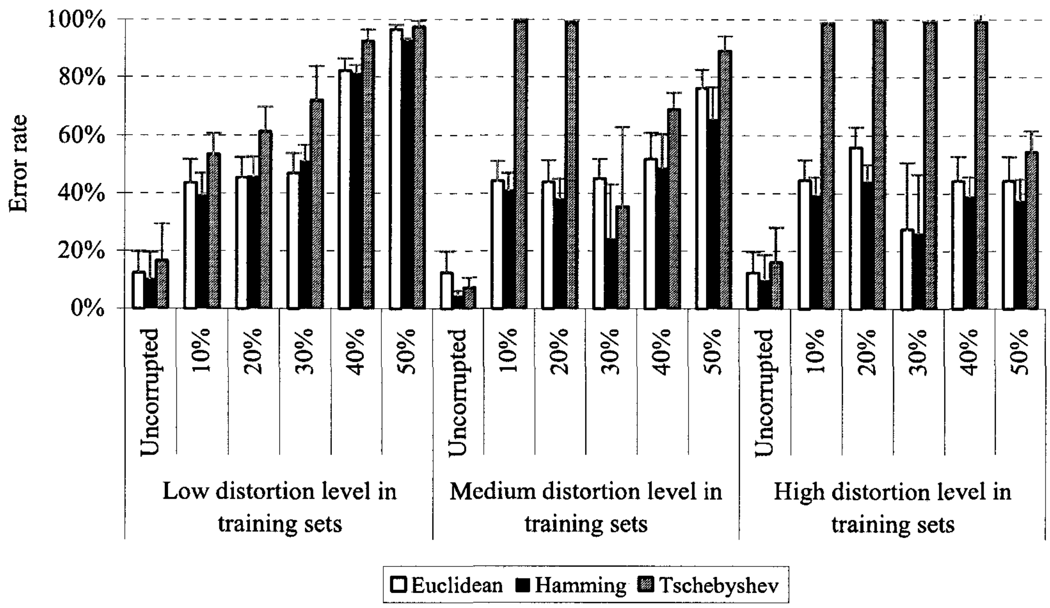


Figure 14 Impact of salt & pepper on Eigenfaces with uncorrupted and corrupted images for training

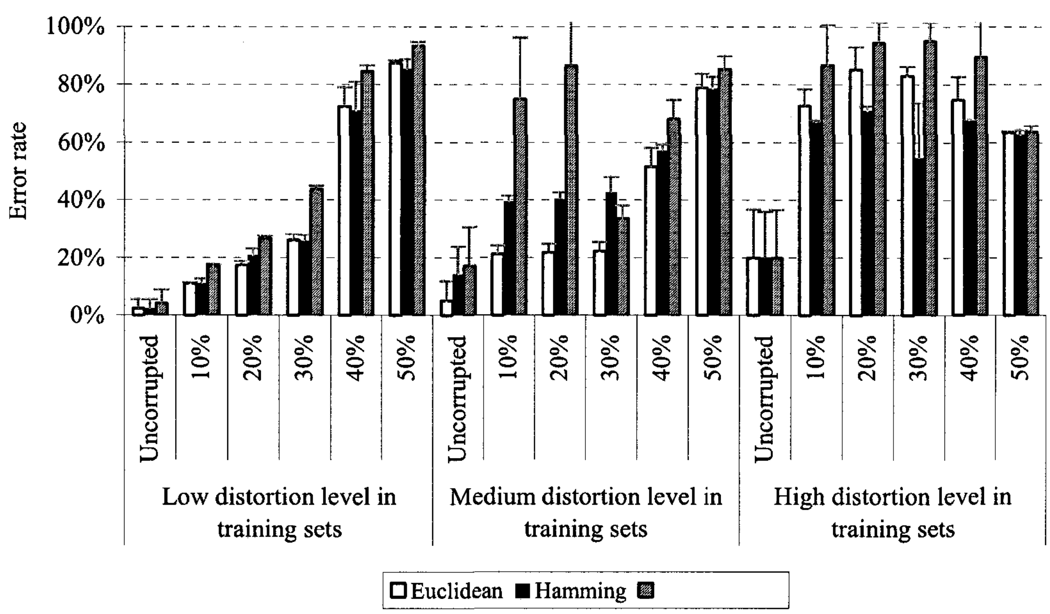


Figure 15 Impact of salt & pepper on Fisherfaces with uncorrupted and corrupted images for training

Lastly Figure 16 and Figure 17 present the impact of Gaussian distortion on Eigenfaces and Fisherfaces methods respectively.

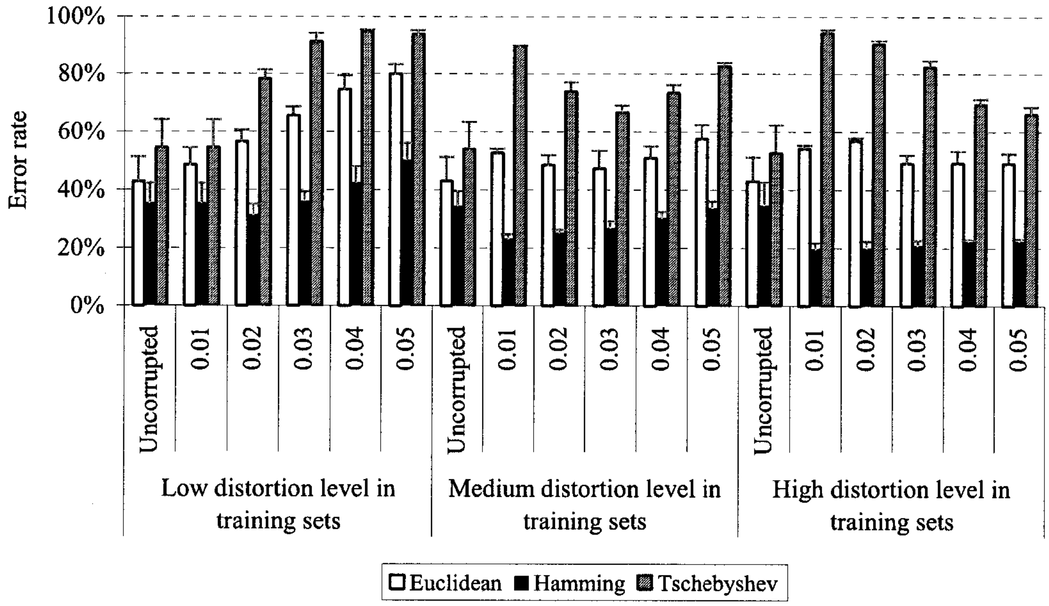


Figure 16 Impact of Gaussian noise on Eigenfaces with uncorrupted and corrupted images for training

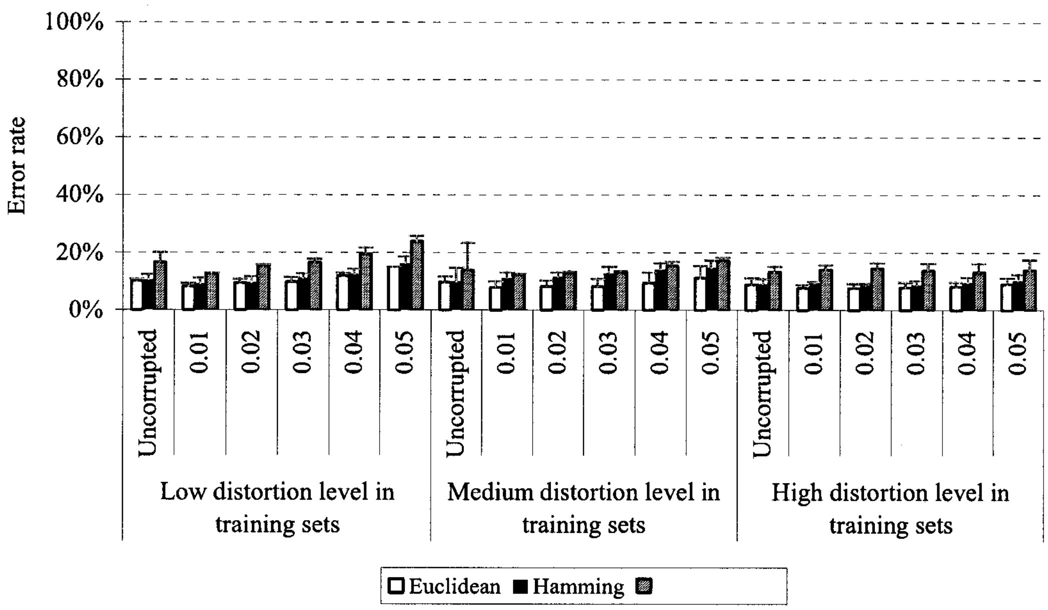


Figure 17 Impact of Gaussian noise on Fisherfaces with uncorrupted and corrupted images for training

From the results it is evident that the Fisherfaces method outperforms the Eigenfaces method for classification tasks. This is expected since the images present differences in lighting conditions and facial expressions. However an interesting thing occurs: the distance functions affect the Eigenfaces and Fisherfaces methods in different manner. In general, Hamming distance provides lower error rates in Eigenfaces face space, and Euclidean distance in Fisherfaces space. Tschebyshev distance is definitely a bad choice in any situation due to the fact that it only considers the largest distance between single dimensions between feature vectors. However it generally follows a trend in accordance with the deterioration level.

Some important hypotheses to consider can be drawn from these findings. Since the face space delineated by Eigenfaces is relatively large, and having in mind that we incur in the small sample size problem, we may find the feature vectors relatively close to each other in such a large space, making the classification task rather difficult. Therefore, a distance measure that separates the feature vectors (or spreads them out along the Eigenfaces space) may help decrease the computed error rates. Let us take a closer look at the Minkowski family of distances described in (1) with a value of $p = 1$ (Hamming distance). We see that the feature vectors may be separated along particular axes (dimensions) of face space when the value of particular features (variables) that form the feature vectors assume negative quantities, i.e. $v_j^{'} < 0$ or $v_j < 0$. Such attribute cannot be pulled off by any other distance used in this study. In the case of Fisherfaces, the face space is considerably narrower than it is in Eigenfaces, therefore using a distance model that separates the feature vectors within the Fisherfaces space may not be as favorable as it is in Eigenfaces space. Let us not forget that Fisherfaces already separates the feature vectors according to classes, and that its transformation matrix D_{LDA} is computed using Euclidean distance implicitly. From this perspective it becomes reasonable that Hamming distance performs better within Eigenfaces space and Euclidean distance within Fisherfaces space. An interesting tendency can also be observed in the impact of blurring on the Eigenfaces method. For low levels of image distortion, in either training and testing sets, the Hamming distance provides lower error rates. However, as the distortion level increases, the Euclidean distance takes over providing lower error rates.

The error rates presented in the scenario 2 show a diminishing tendency when the corruption in the classification sets is similar to the corruption in the training sets. This is not surprising if we think that the feature vectors in the training set already contain the information, in terms if variance, of the corruption in the testing sets. Probably the clearest example can be observed in the error rates shown in Figure 13, where the error rates of training with low distortion level show the lowest values at the point where the classification sets contained low distortion levels as well. On the other hand, the error rates of training with medium distortion level are minimum at the points where testing is performed over uncorrupted and corrupted images with medium distortion levels. Based on our results we suggest introducing image distortion in the training sets to improve classification performance of Eigenfaces and Fisherfaces; there is a significant reduction of the error rate by following this approach.

So far we have introduced only one image distortion model in the training sets at a time. However, it would be of interest to assess the performance of the classifiers when training phase is done with several types of distortion. Such approach may eventually reduce the error rates when classifying images with combined distortion effects. Table 7, Table 8, and Table 9 summarize the findings as to the role of the distance function depending upon the level and intensity of noise. We have identified the type of distance that is favored in the sense it exhibits a tendency of maintaining the lowest error rates.

Table 7 Suggested distances for blurring effect in Eigenfaces and Fisherfaces

Training image set		Uncorrupted	Low distortion	Medium distortion	High distortion
		Testing image set			
Uncorrupted	Eigenfaces	Hamming	Hamming Euclidean	Euclidean	Euclidean
Uncorrupted	Fisherfaces	Euclidean	Euclidean	Euclidean	Euclidean Hamming
Uncorrupted with low distortion	Eigenfaces	Hamming	Hamming	Hamming	Euclidean
Uncorrupted with low distortion	Fisherfaces	Euclidean Hamming	Euclidean Hamming	Euclidean	Euclidean
Uncorrupted with medium distortion	Eigenfaces	Hamming	Hamming	Hamming	Hamming
Uncorrupted with medium distortion	Fisherfaces	Euclidean	Euclidean	Euclidean	Euclidean Hamming
Uncorrupted with high distortion	Eigenfaces	Hamming	Hamming	Hamming	Hamming
Uncorrupted with high distortion	Fisherfaces	Hamming	Euclidean	Hamming	Euclidean Hamming

Table 8 Suggested distances for salt and pepper noise in Eigenfaces and Fisherfaces

Training image set		Uncorrupted	Low distortion	Medium distortion	High distortion
		Testing image set			
Uncorrupted	Eigenfaces	Hamming	Hamming	Hamming	*Any distance
Uncorrupted	Fisherfaces	Euclidean	Euclidean Hamming	Hamming	*Any distance
Uncorrupted with low distortion	Eigenfaces	Hamming	Hamming	Euclidean	*Any distance
Uncorrupted with low distortion	Fisherfaces	Euclidean Hamming	Euclidean Hamming	Euclidean Hamming	*Any distance
Uncorrupted with medium distortion	Eigenfaces	Hamming	Hamming	Hamming	Hamming
Uncorrupted with medium distortion	Fisherfaces	Euclidean	Euclidean	Euclidean	Any distance
Uncorrupted with high distortion	Eigenfaces	Hamming	Hamming	Euclidean Hamming	Hamming
Uncorrupted with high distortion	Fisherfaces	Any distance	Hamming	Hamming	Any distance

* Very high error rate

Table 9 Suggested distances for Gaussian noise in Eigenfaces and Fisherfaces

Training image set		Testing image set: Uncorrupted	Low distortion	Medium distortion	High distortion
Uncorrupted	Eigenfaces	Hamming	Hamming	Hamming	Hamming
	Fisherfaces	Euclidean	Euclidean Hamming	Euclidean Hamming	Euclidean Hamming
Uncorrupted with low distortion	Eigenfaces	Hamming	Hamming	Euclidean	Hamming
	Fisherfaces	Euclidean Hamming	Euclidean Hamming	Euclidean Hamming	Euclidean Hamming
Uncorrupted with medium distortion	Eigenfaces	Hamming	Hamming	Hamming	Hamming
	Fisherfaces	Euclidean	Euclidean	Euclidean	Euclidean
Uncorrupted with high distortion	Eigenfaces	Hamming	Hamming	Hamming	Hamming
	Fisherfaces	Euclidean Hamming	Euclidean Hamming	Euclidean Hamming	Euclidean

Some examples of the misclassified individuals when using the Eigenfaces method are presented in Figure 18, there we show the actual individual being classified along with the individual suggested by the classifier. Likewise the results in Figure 19 concern the misclassified individuals in case of the use of the Fisherfaces method.

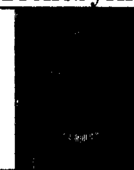
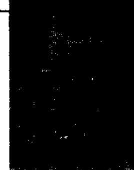
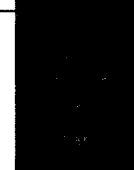
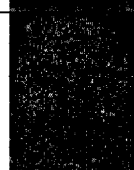
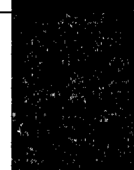
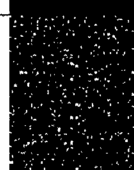
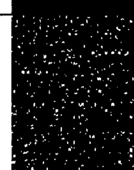
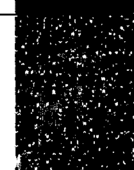
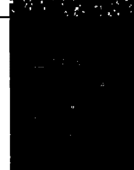
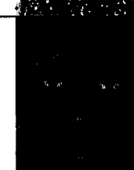
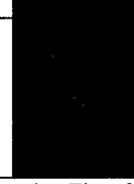
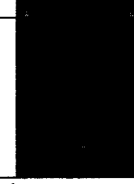
	Distortion	Distance Type Euclidean	 Hamming	 Tschebyshev
Image to be classified	None			
Suggested class	None			
Image to be classified	None			
Suggested class	Gaussian (Variance 0.03)			
Image to be classified	None			
Suggested class	Salt and pepper (Density 10%)			
Image to be classified	None			
Suggested class	Blur (Block 7x7 pixels)			

Figure 18 Example of misclassified individuals using Eigenfaces method

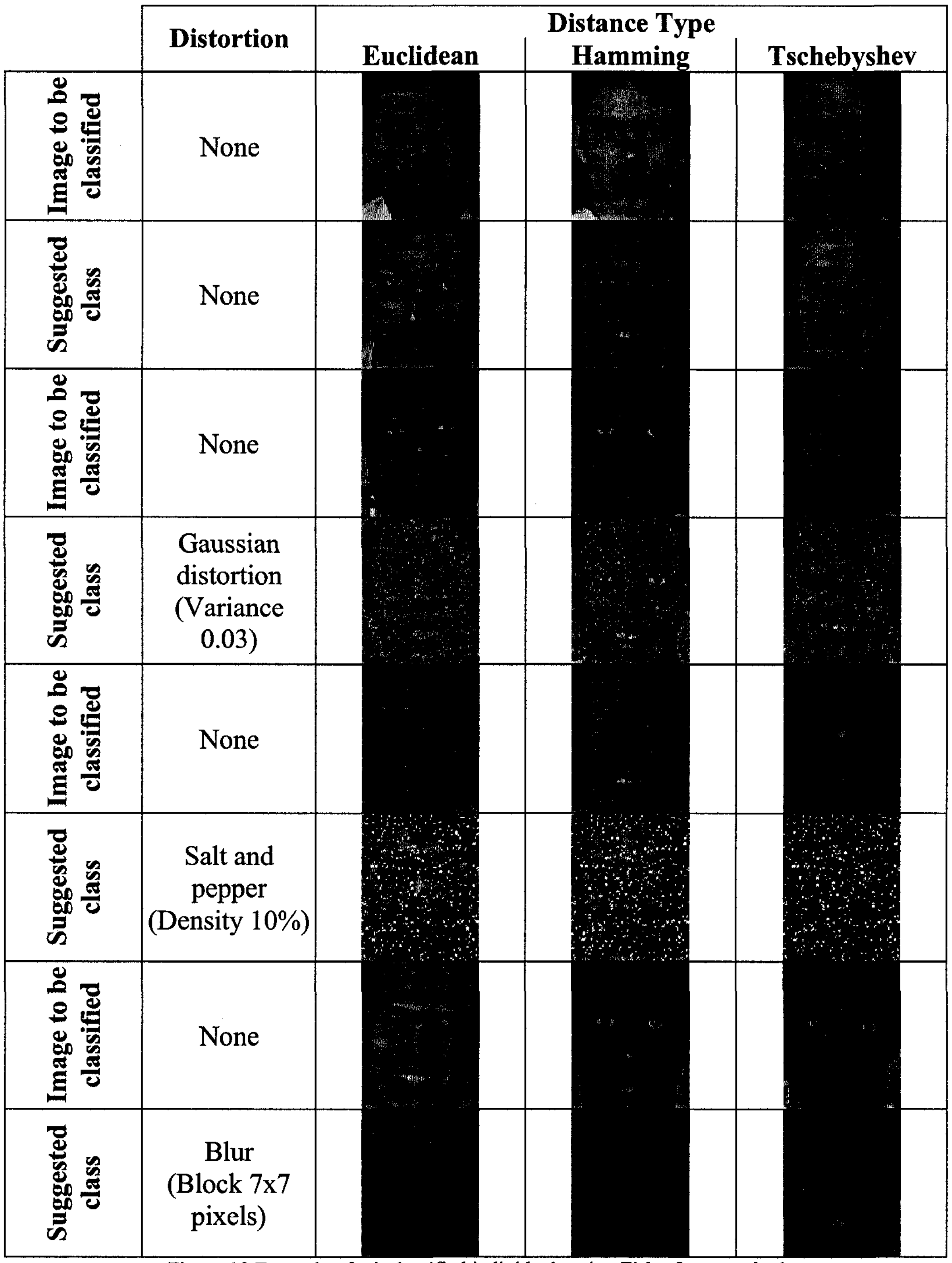


Figure 19 Example of misclassified individuals using Fisherfaces method

5.4 Conclusions

We have delivered an extensive experimental study on the performance of Eigenfaces and Fisherfaces methods completed under noisy conditions. In particular we have investigated blurring effect, salt and pepper, and Gaussian noise. Two general design scenarios have been proposed, they include training with uncorrupted images and training with both, corrupted and uncorrupted images. When designing classifiers, we experimented with three distances that are used to form the nearest neighbor classifier.

We have quantified the effect of noise and arrived at several design guidelines that can be helpful for forming face classifiers in anticipation to various noise conditions and their intensities. In general we have found that the introduction of image distortion in the training sets improves the classification performance of Eigenfaces and Fisherfaces methods. The improvement in classification performance can be attributed, at least partially, to the growth of number of samples in the training sets. In addition, the classifiers could train with distorted images, making them more reliable under noisy conditions. In this study we obtained error rates as low as 1.5%, which as far as we know, has not been accomplished before with the FERET dataset.

Chapter 6

EIGENFACES – A MODULAR APPROACH[2]

Some of the major obstacles that make face recognition technology impractical are attributed to variations of lighting and facial expressions. The literature reports several attempts to overcome such obstacles with the design of numerous approaches. Some attempts involve pre-processing steps, image projection to feature spaces that exhibit robustness against variations of lighting and facial expressions, or classification methods capable of compensating for such variations. Despite evident improvements in face classification, the technology is still susceptible to variations of lighting and gesture.

In this chapter we investigate a technique that intends to overcome the limitations outlined above. We describe a method based on Eigenfaces that follows a modular approach for face recognition, we refer to it as modular Eigenfaces or modular PCA. The modular aspect of the method intends to emphasize particular facial features important for classification, and at the same time, overcome some of the variances attributed to lighting and gesture.

The rationale behind modular PCA finds its revelation on the fact that different portions of a face provide different information in terms of variance, information that is crucial for correct classification. We envision that treating each portion of the face independently could help manage variations of lighting and emphasize important facial features. In modular Eigenfaces, the faces are partitioned into several sections (sub-images) over the horizontal and vertical axes. We extract a total of M sections from each image and use them to generate a new collection of features to represent each individual. Each sub-image is treated independently during the process of feature construction, in this case Eigenfaces.

This chapter presents a comparative study between Eigenfaces and modular Eigenfaces. We evaluate each approach and comment on their advantages and the shortcomings in terms of computational costs, memory requirements, and error rates. We validate our findings on the FERET database as established in section 3.1.

6.1 Previous related work

In 1994, Pentland, *et al.* [77] developed a method based on Eigenfaces that takes into consideration various facial features. A modular eigenspace description technique is used to incorporate salient features such as the eyes, nose and mouth, in an eigenfeature layer. The modular representation led to higher recognition rates as well as a more robust framework for face recognition. They tested their method using a large database comprising approximately 3000 individuals. Cagnoni, *et al.* [13] described a method for face

---

[2] A version of this chapter has been published.
Jarillo, Pedrycz, Reformat, Kwak, "Eigenfaces versus weighted modular PCA for human face recognition", *Proc. of the Symposium on Human-Centric Computing: $HC^2 05$*, February 2005, Banff, Alberta, Canada.

recognition based on the eigenimage technique. They extracted the regions of the eyes, nose, and mouth and performed dimensionality reduction via PCA on each region. The features are compared by nearest neighbor classifier and Euclidean distance. They tested the method using 144 images from the Vision and Modeling Group at the MIT Media Lab, and 112 images from the Department of Computer Science and Applied Mathematics of the University of Berne. The authors reported classification rates of about 95% with the modular approach. Melin, *et al.* [71] published their work in face and fingerprint recognition, their work is based on the concept of modular approach. Their system divides a human face into three different regions, namely eyes, the nose and the mouth. Each region is assigned to one module of a neural network. Hence, the modular neural network has three different modules, one for each region. The final decision is computed by an integration module, which takes into consideration the results of each module. The integration uses the fuzzy Sugeno integral to combine the outputs. They tested their method using their own face database. Liao, *et al.* [57] developed an automatic face recognition system based on multiple facial features. Each facial feature is represented by a Gabor-based complex vector and is localized by an automatic facial feature detection scheme. They proposed two approaches for recognition, named Two-Layer Nearest Neighbor (TLNN) and Modular Nearest Feature Line (MNFL). They validated their findings using the Cambridge, YALE, and Harvard databases. They report recognition rates of about 96%.

6.2 Modular PCA

The scheme behind modular PCA emerges from the idea that distinct sections of a face provide different information in terms of variance. Aspects such as lighting can be diminished if the images are divided into various regions, hence PCA would produce basis vectors with local variances of each region. That information may prove crucial for the correct overall classification. As already pointed out, the faces are partitioned into several sub-images over the horizontal and vertical axes, and all sub-images are of the same size. We extract a total of M portions to form a new collection of images to represent each individual. It is expected that most of the variance exists mainly on some distinctive features of the face, creating in this way a pattern that provides useful information from the areas that are most important for classification. Since the new collection is proportionally larger to the number of sub-images M, the computational cost becomes an issue to consider.

Figure 20 illustrates the concept behind modular PCA [38]. The images are uniformly segmented into distinctive sections. The newly produced sub-images form the new gallery set. Each set of sub-images is transformed independently by PCA, which produces an independent set of features. At this point it is important to keep in mind the size of the feature vectors as they will form a larger representative feature vector (M times larger than in Eigenfaces). The feature vectors are combined in the feature space, one after the other, to form a new representative feature vector. Once the new representative feature vectors are formed, it is possible to emphasize specific areas of the face by using weights for each region. This can be achieved by using a weighted distance during classification.

In modular PCA the number of extracted sub-images brings two issues to consider. The first refers to the computational cost, as it escalates as the number of extracted sub-images increases. The second relates to the storage requirements for the representative features, as it enlarges in the same manner.

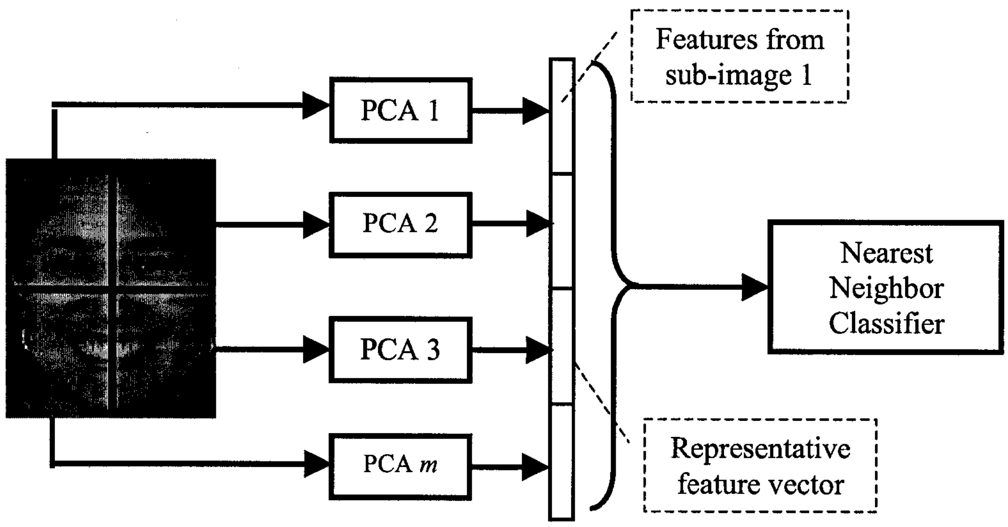


Figure 20 Modular PCA scheme

In this analysis we used weighted distance for the nearest neighbor classifier in order to set specific "importance" to each section of the face. The weights are computed according to the portion of the variance that each section accounts for in relation to the overall face. The following correspondence is used

$$u_s = \frac{\sum_{\lambda_l \in s} \lambda_l}{\sum_{j=1}^{Q} \lambda} \tag{20}$$

where u_s, $s = 1,...,M$, are weights associated to the region s, and λ_l, $l = 1,\ldots,N$, are the eigenvalues of the covariance matrix formed by the sub-images of region s. The vector $\mathbf{u} = [u_1,..., u_M]^T$ defines the "importance" of each portion of the face. Q stands for the total number of eigenvectors involved, i.e. M x S. The weighted distance is computed taking into account the representative feature vectors modified by the weights in $\mathbf{u}$.

6.3 Research environment

The research environment uses images of the FERET database. The training and testing sets comprise 400 and 200 images respectively. Since there are 3 images per person, a 3-fold cross validation is performed to achieve an unbiased behavior of both methods. Each gallery set is obtained randomly, 2 images for training and the remaining one for testing, with no overlap between gallery sets. In this study, the images show only the face area, the images comprise 80x100 pixels by 8 bits per pixel, as presented in section 3.3.

6.4 Experimental setup

In this comparative study we evaluate the performance of Eigenfaces and modular PCA. The intention is to quantify the performance of both classifiers in order to assess their robustness against variations of lighting and expressions, we expect modular PCA to outperform PCA. Eigenfaces serves as baseline for our investigation since it is a method known to have limitations against lighting and facial expressions. The performance of the classifiers is expressed in terms of error rates.

We depart from the assumption that every section of the face provides, to some extent, different distinctiveness among images, information that is relevant for the correct classification. For instance, the cheek may not be a very distinctive feature among the individuals, and therefore providing lower variance compared to other regions of the face. It is palpable that Eigenfaces does not acquire information regarding the distinctiveness between regions of the faces, even though it reveals the variance among images forming the covariance matrix.

Given that it is unknown which regions of the face provide the highest variances "distinctiveness", we systematically divide the images over the horizontal and vertical axes in order to test the modular approach. The images are divided from 1 to 5 sections over each axis, making an extraction of up to 25 sub-images from every face. The final representative feature vector is formed by combining the feature vectors obtained from all section independently transformed by PCA, one vector after another. Each section is then weighted according to (20) in order to set its importance. At the end of the tests, the error rates delivered by modular PCA are compared to those of Eigenfaces. General guidelines for proper subdivision of the images are discussed.

The construction of the feature space by means of the modular approach produces feature vectors of 400 variables to represent each sub-image. However we only retain 100 variables. The retained cumulative variance accounts for approximately 96% to 98% of the total variance. In our work published in 2006 [39] we found that the performance of the classifier stabilizes after retaining 90% of the total variance. We test modular PCA by taking into consideration various numbers of variables. The amount that provides the lowest error rate is presented along with the corresponding classification performance.

In Eigenfaces, the feature space comprises 400 dimensions, the performance of Eigenfaces is tested taking into account from one to 400 variables. The number of variables that deliver the lowest error rate is reported.

6.5 Experimental results

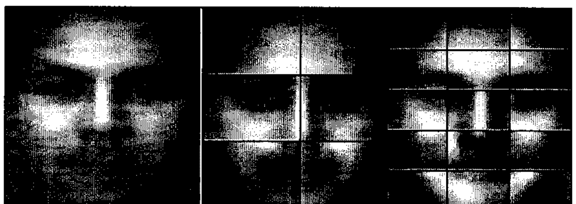
The reported data is expressed in terms of average error rate over the testing sets. In all cases the feature vectors are formed taking into account 100 eigenvectors of each sub-eigenspace, they are selected according to the highest variances.

Eigenfaces had an error rate of 36.5%. In contrast, Table 10 presents the experimental results (expressed as percentage error rates) obtained for modular PCA. The data is organized according to the sub-divisions made along each axis of the images. The lowest and highest error rates are indicated.

Table 10 Error rates of modular PCA

		Number of sub-divisions along the horizontal axis				
		1	**2**	**3**	**4**	**5**
Number of sub-divisions along the vertical axis	**1**		28.5	29.0	25.0	27.5
	2	26.5	27.5	25.0	27.0	29.0
	3	25.5	**24.5**	25.5	25.5	28.5
	4	27.0	28.0	28.5	29.5	29.0
	5	25.5	28.5	**31.5**	29.5	31.0

Figure 21 depicts the average images obtained from Eigenfaces and modular PCA. The figure displays the average sub-images for the configurations that provide the lowest and highest error rates in modular PCA in accordance to Table 10.



Traditional Eigenfaces	Modular PCA (2 x 3)	Modular PCA (3 x 5)

Figure 21 Average images of Eigenfaces and modular PCA

The images presented in Figure 21 may look like they do not match perfectly, the reason is that the pixels' values are not integer numbers during computations; consequently an error is introduced by the rounding off process. In addition, the values were rescaled from 0 to 255 to generate the images. However, the true average values were considered throughout the computations.

Table 11 shows the weights associated with each section of the face for the cases where the error rates assume their lowest and highest values. We note that the regions of the face that provide most of the variance with respect to others are the top and bottom. Furthermore, the left side of the images seems to provide better distinctiveness for classification, possibly due to particular lighting conditions portrayed in the dataset.

Table 11 Weights associated with specific images regions (cases resulting in lowest and highest error rates)

0.2158	0.1754
0.1500	0.1194
0.1907	0.1484

Modular PCA (2 x 3)

0.0881	0.1044	0.0652
0.0705	0.0614	0.0587
0.0627	0.0478	0.0484
0.0742	0.0591	0.0519
0.0853	0.0592	0.0623

Modular PCA (3 x 5)

6.6 Conclusions

Our findings suggest that modular PCA is a valid approach for improving classification of human faces over Eigenfaces. We report lower error rates with modular PCA than with Eigenfaces method, as much as 12% for the best case, and 5% for the worst. We have also found that, as a general trend, the upper and lower regions of the faces are most distinctive for classification according to the portion of the variance they account for. Perhaps physical aspects such as hairstyle provide distinctiveness for classification. Other weighting factors, distance models, and classifiers may be considered in both approaches, which we admit may produce better performances. However, this study focuses on the fair comparison between Eigenfaces and modular approaches only.

Chapter 7

IMPACT OF IMAGE RESOLUTION IN FACE CLASSIFIERS[3]

The current literature presents an overwhelming collection of experimental investigation in the subject of face recognition. There is a constant flow of ideas, methodologies and architectures being proposed and investigated in the field. Fair assessments and comparisons of reported investigations are crucial for the improvement of face recognition technology. To this end, well-established research methodologies and research environments have been proposed and widely adopted by the scientific community. For instance, current research methodologies and protocols involve manners in which classification accuracy is measured and reported. In face recognition one can evaluate a system's performance in various well-defined ways depending on the intended purpose of the system. If a system is to single out the identity of an individual from a collection of persons, then the performance is commonly expressed in terms of error rates. If the system is to operate in environments where access to resources is restricted to all but few individuals, then the system's performance is expressed in terms of False Acceptance Rates (FAR) and False Rejection Rates (FRR).

In terms of the research environments, several databases regarded as benchmarks have been proposed and made available for researchers to experiment with. Such databases are, without question, priceless contributions that enable researchers to make fair assessments of proposed systems. Nonetheless, it is still sometimes difficult to evaluate and compare various methodologies proposed by independent researchers. It is common to find differences in reported results and derived conclusions in the literature, even when the methodologies and research environments were kept the same by the researchers. Some causes for such differences may include discrepancies in the data taken into consideration. Even when most researchers adopt the same databases, issues such as image quality and pre-processing steps can lead to disagreements in the findings.

Regarding image quality, researchers commonly manipulate the original databases in order to make experimentation feasible. Due to the highly computational nature of many algorithms in face recognition, and in pattern recognition in general, the databases are commonly simplified. In face recognition, downsampling of the original patterns is a typical approach to reduce computational cost for many algorithms. However, the loss of information can carry negative implications in the system's performance. As of today, there is not a clear understanding as to what is the adequate or minimum image resolution to adopt while preserving accuracy. The literature does not offer much information taking into account such a tradeoff, hence an investigation is required.

---

[3] A version of this chapter has been published.
Jarillo, Pedrycz, Reformat, "Analysis of image resolution and transformation in face classifiers", *Proc. of the Symposium on Human-Centric Computing and Data Processing: HC^2DP_{07}*, February 2007, Banff, Alberta, Canada.

This chapter offers a comprehensive investigation aiming at revealing and quantifying the performance of face classifiers under various image resolutions (image qualities). To extend the reach of our investigation we include various image transformations to construct our databases. Particularly we take into consideration transformations such as contrast enhancement and edge detection. The selected image transformations are frequently included as pre-processing steps in many face recognition systems, let it be for enhancement purposes or extraction of facial features. Within the context of face classifiers, we adopt methods capable of sound dimensionality reduction, namely Eigenfaces, Fisherfaces, kernel-PCA, and Isomap. For details on each method please refer to Chapter 4.

Our findings portray practical implications to system's design and performance. The fact that this study takes into consideration several methods for dimensionality reduction, image transformations, and image resolutions make this contribution valuable and unique.

7.1 Previous related work

The fact is that the literature does not offer much material to make an extensive literature review regarding the impact of image resolution over face recognition algorithms. Nonetheless we can still highlight interesting investigation carried out in the past. Anjum, *et al.* [6] reported their investigation on the effect of image resolution on a linear method of dimensionality reduction on images depicted by pixels. They did their evaluations on the ORL, YALE, FERET, and EME databases. They established that for each database there is always an optimal image resolution where the recognition performance is best. Kouzani, *et al*, [49] presented another investigation involving variations of image resolutions. They proposed a system that assigns different degrees of importance to each part of a face, each region is processed with a different resolution. Their system reduces the computational complexity and achieves higher recognition rates in comparison with the Eigenfaces method. They validate their finding on their own database consisting of 200 individuals. They report 100% recognition rate using their proposed method. McLindin, *et al.* [70] quantified the effects of using different types and resolutions of gallery images on two different commercial face recognition systems, namely Face-It and Argus 2D. They experimented using a their own dataset obtained from operational trials. They concluded that images with more than 30 pixels of separation between the eyes are suitable for recognition. More recently, Arca *et al.* [4] measured the effect of image resolution and gallery size in face recognition. Their approach is described in [3]. They did their study on the FRGC 1.0 and the XM2VTS databases. They varied the image scale, so that the intra-ocular distance changed from 50 to 250 pixels with a step of 25. They observed that the behavior does not change significantly within the range of 75 – 250 pixels, while if the intra-ocular distance is lower than 75 pixels the performance decreases.

7.2 Research environment

The research environment allows us to reveal and quantify the relationship that exists between image quality (resolution) and classifier's performance. The research environment comprehends a collection of face databases and image transformations. Various levels of image resolutions are taken into consideration in this study. The quality of the images is deteriorated by means of average downsampling. The databases taken into consideration for this study include FERET and YALE, they have been described in section 3.3. The classifiers under discussion include linear and non-linear methods, such as Eigenfaces, Fisherfaces, Kernel-PCA, and Isomap. Details on the face classifiers can be found in Chapter 4.

We consider important to include image transformations as part of this investigation. Image transformations have been adopted and presented in the literature of face recognition, for examples refer to [94][101]. In this investigation we focus on contrast enhancement by means of histogram equalization and on edge detection by the Sobel operator. The main purpose of histogram equalization is to compensate for variations of lighting and to reduce the influence of shadows in face classifiers. Edge detection is generally implemented to extract facial features useful for classification, such as eyes, nose, moth, and contour of the face. Contrast enhancement and edge detection are explained in section 3.3.

A number of downsampling factors are considered in this investigation. They lead to typical image resolutions presented in the literature of face recognition. For the FERET database we include downsampling by factors of 2, 5, 7, 8, and 9; leading to image resolutions of 80x100, 32x40, 22x28, 20x25, and 17x22 pixels respectively. For the YALE database we included no downsampling as well as downsampling factors of 2, 3, 6, and 9, leading to image resolutions of 144x150, 72x75, 48x50, 24x25, and 16x16 pixels respectively. The process of downsampling is explained in the next section.

7.2.1 Image downsampling

The downsampling process adopted in this study reduces the size of an image by representing a group of neighboring pixels by the average of their values. A scaling factor sets the number of pixels in the vertical and horizontal axis required to compute the average values. The computed average is rounded off to the closest integer in order to represent a gray level. The scaling factor is set by the user as an integer number. Figure 22 illustrates an example of downsampling with a sampling factor of 2. Each number in the matrix represents a pixel value. In the example, a region of size 2x2 is taken into consideration to compute the average. During the downsampling process, the region (window) slides along the horizontal and vertical directions of the entire image.

$$\begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 4 & 5 & 6 \\ \hline 7 & 8 & 9 \\ \hline \end{array} \left.\vphantom{\begin{array}{c}1\\4\\7\end{array}}\right\} \text{Average gray level} = 7$$

Figure 22 Example of downsampling

Figure 23 and Figure 24 show some images of the FERET and YALE databases respectively with variations in image quality (resolution). The figures depict images with no transformation, contrast enhancement, and edge detection. The images with higher resolution in Figure 24 were scaled down to fit the page. The leftmost column indicates the actual image size in terms of horizontal and vertical pixels.

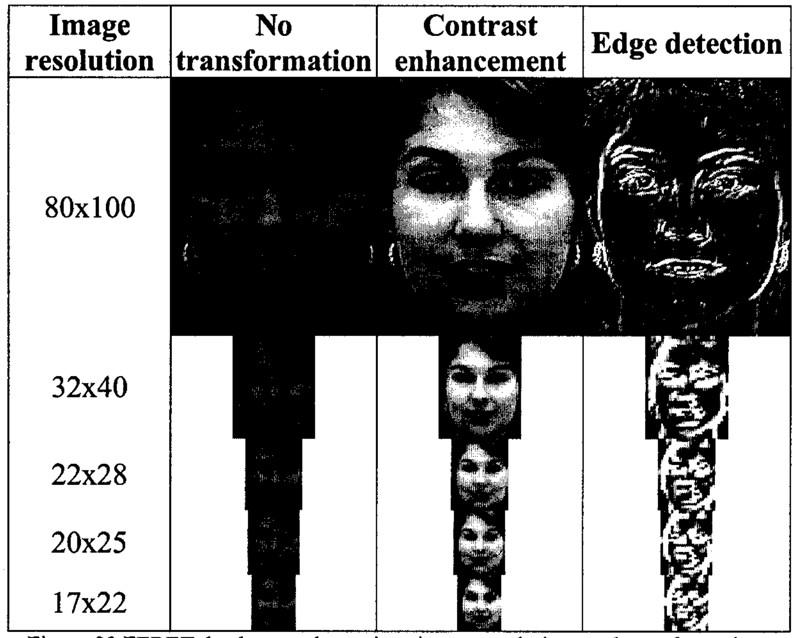


Figure 23 FERET database under various image resolutions and transformations

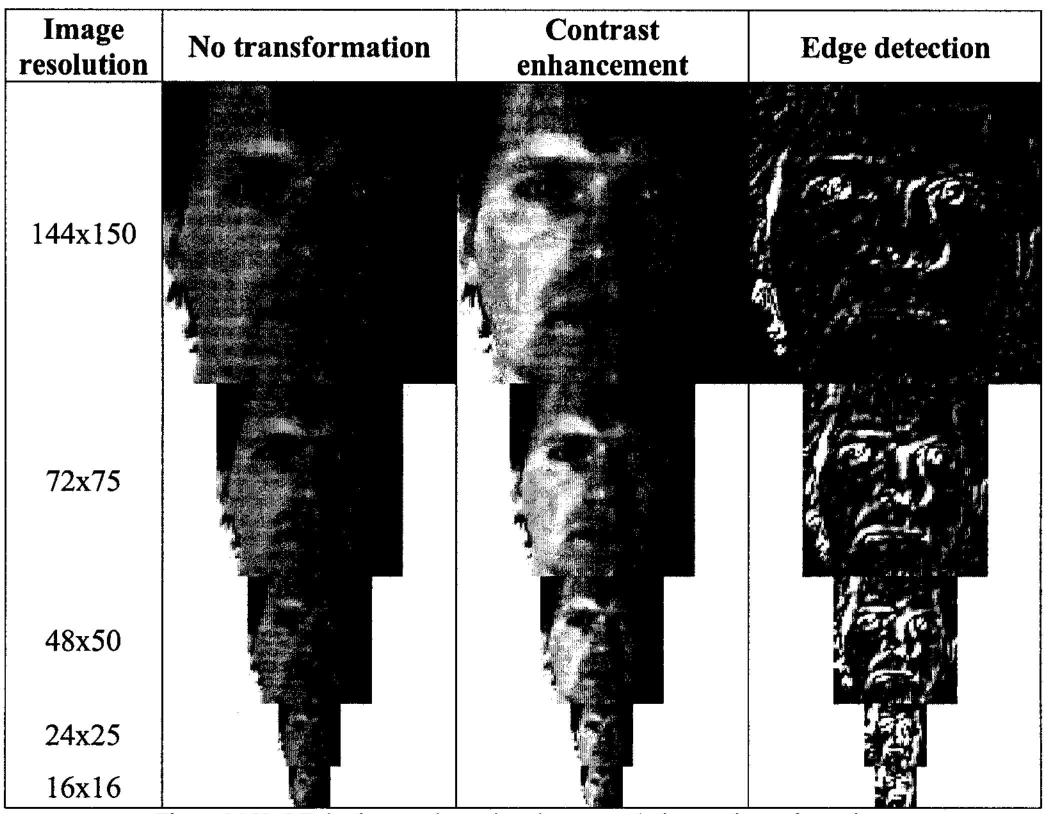


Figure 24 YALE database under various image resolutions and transformations

7.3 Experimental setup

The experimental setup is designed to reveal the relationship that exists between image quality, portrayed as image resolution, and classifier performance, expressed in terms of error rates. We allow two image transformations in this investigation: contrast enhancement carried out by histogram equalization, and edge detection by means of the Sobel operator. To improve the performance of the edge detection process, the images were pre-processed by contrast enhancement. The Experimentation is carried out using the FERET and YALE databases. Both databases are pre-processed to generate two more categories of databases, namely contrast-enhanced images and edge images. Each category is treated independently throughout the entire experimentation.

There are three images per person in the FERET database for a total of 200 individuals. The FERET images were divided into 400 images for training, and 200 for testing with no overlap (one image per class in each set). A 3-fold cross validation is taken into account to evaluate those architectures using the FERET database.

The YALE database contains 11 images per person for a total of 15 individuals. We created 10 random splits with overlap, each including one training and one testing set comprising 6 and 5 images per person respectively. The evaluation of the architectures using the YALE datasets is reported as the average and standard deviation computed over the 10 splits.

Various feature spaces are taken into consideration in out investigation, they include linear and non-linear methods of sound dimensionality reduction, such as Eigenfaces, Fisherfaces, Kernel-PCA, and Isomap. The proposed feature spaces allow for a variety of configurations, such configurations may involve specific parameters for the construction particular face spaces or for the selection of adequate variables for classification. A careful selection of these parameters is essential in order to achieve adequate performance of the classifiers.

In the proposed classifiers it is convenient play with the number of variables involved in the classification process. Usually the variances along each axis of the feature space are considered in order to select the variables with most discriminatory power. Still it is necessary to experiment with a number of variables in order to select those that produce the lowest error rates. In this study we tested the proposed classifiers taking into account variables chosen according to their variances. The number of variables that led to the lowest error rate was then selected as proper for the given classifier. In this study, we report only the lowest error rates we computed considering a particular space and number of variables.

Some algorithms for the construction of feature spaces are flexible enough to allow researchers specify some configurations. In kernel-PCA, the user is to decide the kernel model and the parameters for it. A complete description of kernel-PCA and its parameters is presented in section 4.4. In this study we investigated three kernel models, namely polynomial, Gaussian, and sigmoid. The parameters assumed by each kernel are presented in Table 12.

Table 12 Kernel parameters for kernel-PCA

Kernel model	FERET	YALE
Polynomial	d = 2, 3, 4, and 5.	d = 2, 3, 4, and 5.
Gaussian	σ = 10, 30, 50, 70, 90, 100, and 150.	σ = 100, 150, 200, and 250.
Sigmoid	Polynomial kernel with d = 1, 2, 3, and 4.	Polynomial kernel with d = 1, 2, 3, and 4.

The Isomap algorithm requires the construction of the matrix of pairwise geodesic distances D. For a complete description of the algorithms and parameter of Isomap please refer to section 4.5. D is constructed by connecting neighboring data samples (feature vectors) in a given space. One can connect each sample either by following the k-nearest neighbor (k-NN) algorithm, which requires a value for parameter K, or by specifying a radius ε to enclose neighboring samples. If K is equal to the number of images, then we end up computing regular PCA, if K is too large we may not achieve any reduction of the variables. Hence the goal is to find a small value of K capable of connecting all data samples. In this study we connected the feature vector by following the k-nearest neighbor algorithm. We iteratively ran Isomap with increasing values of K. We started the iterations with $K = 1$, if one or more elements of D were not connected at the end of the process then K was incremented by 1 and the k-NN algorithm started again. Evidently if K assumes the value N we end up computing regular PCA.

The Isomap algorithm is computationally expensive, the computation of geodesic distances involves computing the K nearest neighbors as to connect all samples, followed by Dijkstra's algorithm to find the geodesics. For such reason, using face images (depicted by n pixels) as inputs for Isomap is impractical. Instead we reduced the number of variables by constructing Eigenfaces and Fisherfaces spaces prior to implementing Isomap. Hence we have two scenarios for Isomap: one including Eigenfaces features and the second counting Fisherfaces features. In both cases the entire feature spaces were taken into account in Isomap.

The assessment of the classifier performance with respect to image quality (resolutions) and image transformations allows us to draw design guidelines for face recognition systems. The assessment is delivered in terms of error rates and corresponding standard deviations. It also reveals the number of variables that were required to achieve the reported error rates.

7.4 Experimental results

The assessment of the classifier performance is presented in terms error rates and corresponding standard deviations. For the sake of clarity we report only the lowest error rates that we computed within the collection of experiments. The complete suite of experiments considered variations in the parameters in each algorithm, as explained in the previous experimental setup.

The performance of the classifiers is presented in a variety of charts. The evaluation of Eigenfaces and Fisherfaces is depicted in Figure 25. Figure 26 shows the performance of the Isomap algorithm for both scenarios: Eigen-Isomap and Fisher-Isomap. To complete, Figure 27 portrays the performance of kernel-PCA taking into account three kernel models, namely polynomial, Gaussian, and sigmoid. The charts include the performance of the classifiers taking into account images with no transformation, contrast-enhanced images, and edge images using the YALE database. The data is organized according to feature space and image quality. The error rates are depicted by bars in each instance. The standard deviations are portrayed as thin lines overlapped with the error rates.

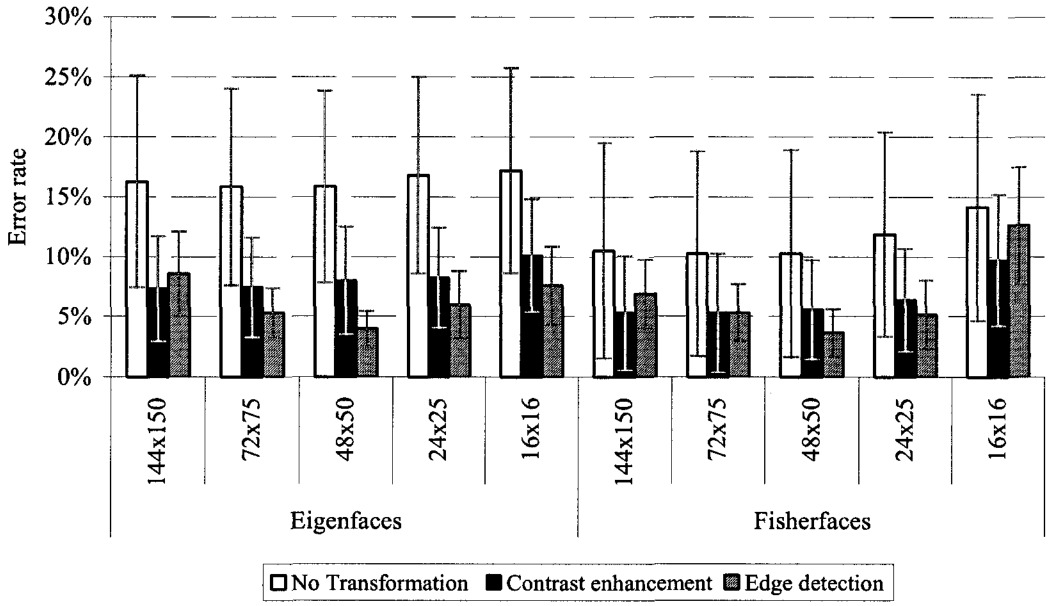


Figure 25 Performance of Eigenfaces and Fisherfaces using YALE

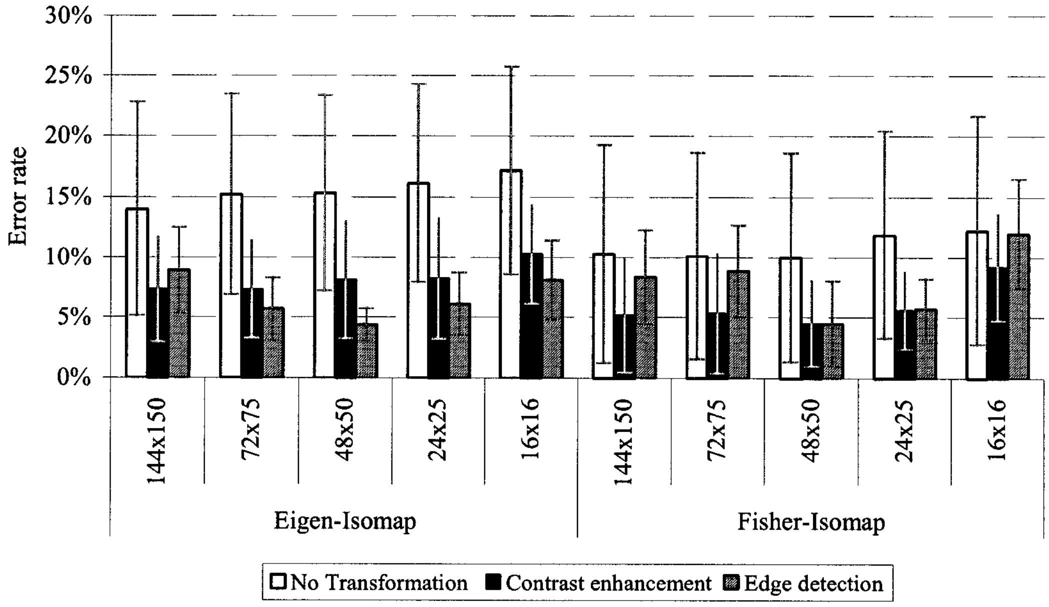


Figure 26 Performance of Eigen-Isomap and Fisher-Isomap using YALE

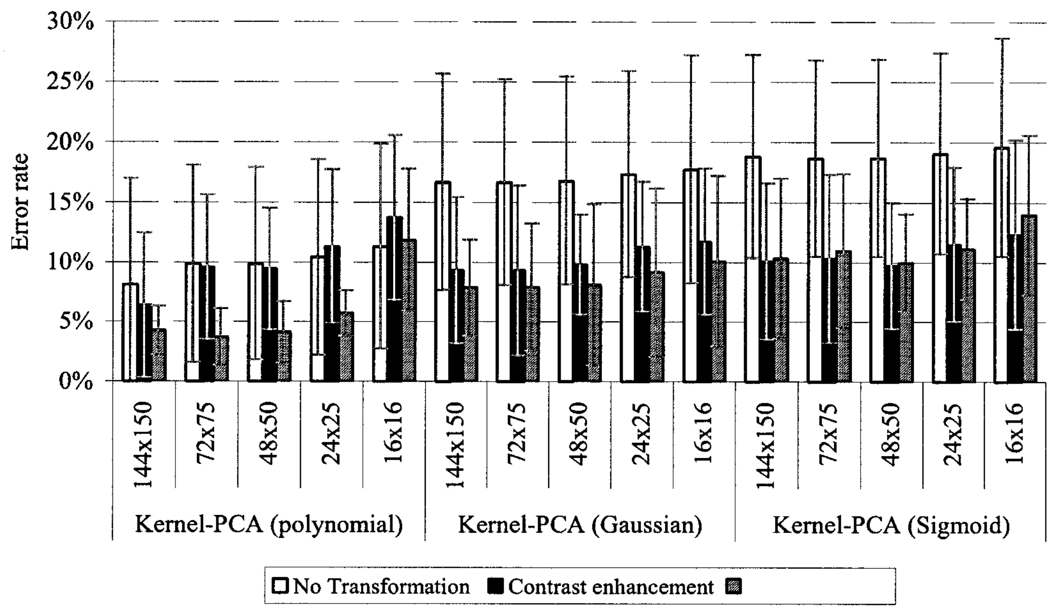


Figure 27 Performance of kernel-PCA using YALE

Table 13 shows the parameters of each classifier and the number of variables that led to the lowest error rates presented above. Note that for the Isomap algorithms we report a range of values for K. We computed the error rates over 10 random splits of the YALE database, hence each split has its own value of K. We report the ranges that include all values of K assumed by the 10 splits.

Table 13 Parameter settings for each feature space using YALE

Method	No transformation		Contrast enhancement		Edge detection	
	Parameter(s)	Variables	Parameter(s)	Variables	Parameter(s)	Variables
PCA	NONE	90	NONE	90	NONE	85
LDA	NONE	14	NONE	14	NONE	14
Eigen-Isomap	$5 \leq K \leq 10$	75	$9 \leq K \leq 13$	90	$12 \leq K \leq 19$	90
Fisher-Isomap	$6 \leq K \leq 12$	90	$9 \leq K \leq 14$	75	$11 \leq K \leq 17$	75
Kernel-PCA (polynomial)	d=2	85	d=2	59	d=2	51
Kernel-PCA (Gaussian)	σ=200	27	σ=200	48	σ=200	48
Kernel-PCA (Sigmoid)	Polynomial kernel (d=2)	46	Polynomial kernel (d=1)	48	Polynomial kernel (d=1)	36

The performance of the classifiers taking into account images of the FERET database is presented as follows: Figure 28 shows the performance of Eigenfaces and Fisherfaces, Figure 29 depicts the error rates of Isomap for both scenarios, and Figure 30 portrays the results of kernel-PCA.

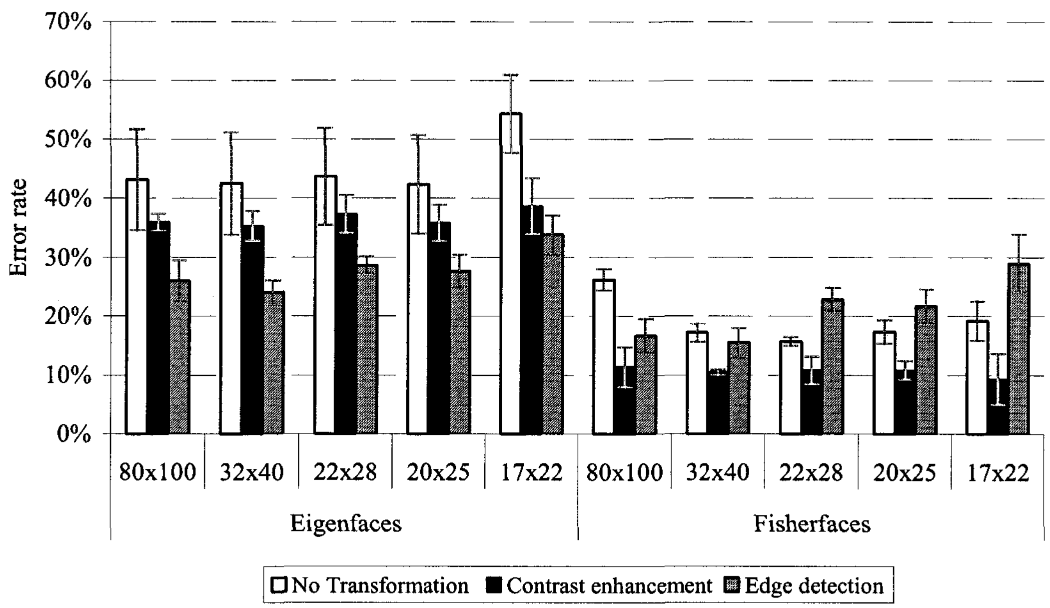


Figure 28 Performance of Eigenfaces and Fisherfaces using FERET

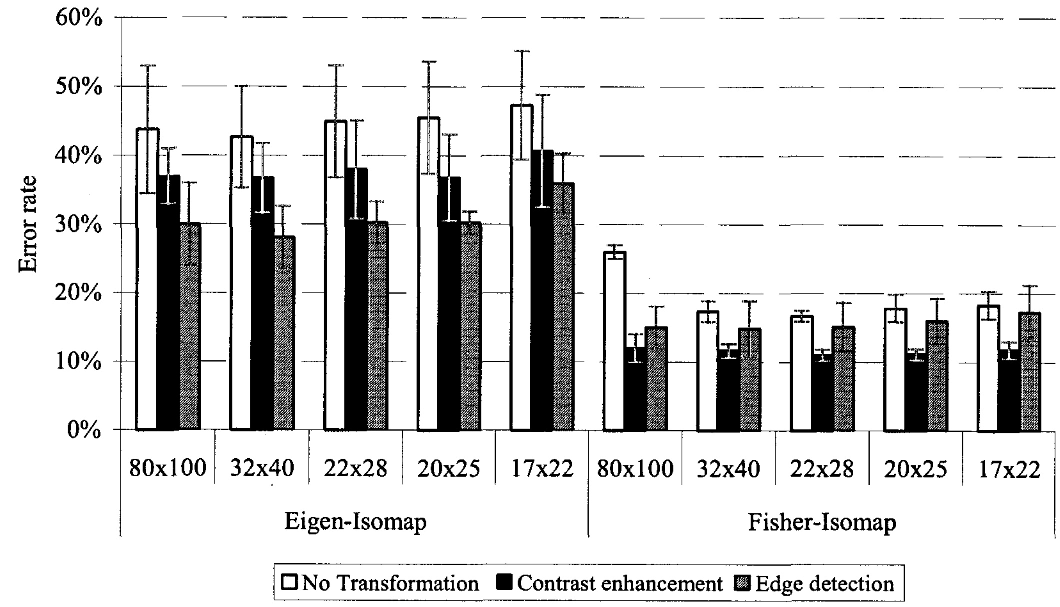


Figure 29 Performance of Eigen-Isomap and Fisher-Isomap using FERET

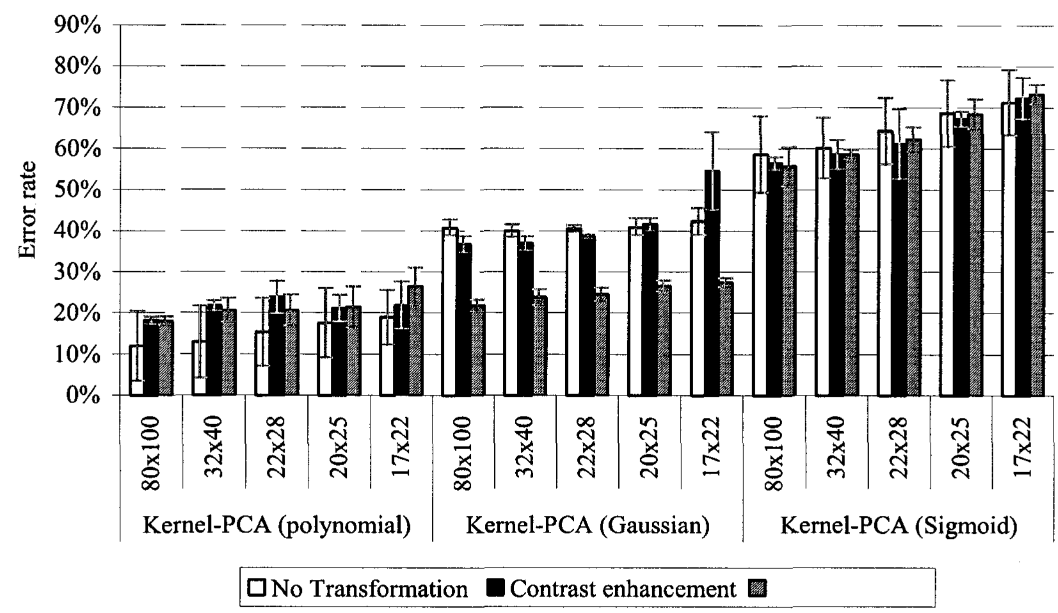


Figure 30 Performance of kernel-PCA using FERET

Table 14 shows the parameters of each classifier and the number of variables that led to the lowest error rates using the FERET database. Once more, note that for the Isomap algorithms we report a range of values for *K*. The error rates over the FERET database were computed by a three-fold cross validation, hence each validation split has its own value of *K*. We report the ranges that include all values of *K* for the three splits.

Table 14 Parameter settings for each feature space using FERET

Method	No transformation		Contrast enhancement		Edge detection	
	Parameter(s)	Variables	Parameter(s)	Variables	Parameter(s)	Variables
PCA	NONE	180	NONE	235	NONE	228
LDA	NONE	70	NONE	128	NONE	123
Eigen-Isomap	$39 \le K \le 48$	384	$59 \le K \le 68$	399	$64 \le K \le 78$	398
Fisher-Isomap	$27 \le K \le 39$	199	$47 \le K \le 59$	182	$51 \le K \le 57$	185
Kernel-PCA (polynomial)	$d=2$	384	$d=2$	190	$d=2$	174
Kernel-PCA (Gaussian)	$\sigma=70$	398	$\sigma=70$	355	$\sigma=70$	374
Kernel-PCA (Sigmoid)	Polynomial kernel ($d=1$)	124	Polynomial kernel ($d=1$)	132	Polynomial kernel ($d=1$)	160

For the sake of clarity, this section reports only the lowest error rates that we were able to compute from an extensive variety of experiments. The complete suite of experiments includes a number of variations in the parameters required by each classifier. It also includes a variety of image resolutions, and face images with and without transformations.

In section 3.1, we introduced two different image distributions (scenarios) for the evaluation of the face classifiers:

- *Scenario 1*: involves training and testing sets.
- *Scenario 2*: includes training, validation, and testing sets.

Table 1 provides details on the image distributions in each case. The experimental results presented in this section were computed taking into consideration the training and testing sets of the FERET and YALE databases (scenario 1). So far we included 2 images for training and 1 image for testing in the experimentation with FERET; and 6 images for training and 5 for testing in the experimentation with YALE. In Appendix B we provide the experimental results taking into consideration the image distributions of scenario 2. The feature spaces were constructed with the images of the training sets and the classification was performed over the testing sets. The validation set is not required for the investigations reported in this chapter. The data presented in Appendix B is useful for comparison purposes with other investigations reported in this thesis.

Based on the provided experimental results we can appreciate the effect image resolution in face recognition. The relationship existing between error rates vis-à-vis image resolution has been revealed for a number of classifiers and image transformations. Overall, the classifiers' performance tends to deteriorate as we reduce the resolution of the images. However we can see that there are particular image resolutions that allow better classification, particularly when the classifiers deal with images without any transformation. For instance, in the case of Fisherfaces and Fisher-Isomap the error rates drop to the lowest point with image resolutions of 32x40 and 22x28 respectively, see Figure 28 and Figure 29. This trend is appreciated in both databases and various classifiers. Allowing some loose of generality, the image resolutions that deliver the lowest error rates are 22x28 when using YALE, and 72x75 in the case of FERET. We initially expected to show deterioration in classification performance with respect to image degradation. However, to our surprise the performance of the classifiers is not better when they consider image with higher resolution. The experimental results suggest that problem of dimensionality reduction (reducing the number of variables of the images represented by pixels) is simplified when we consider less amount of data from the input space. This is an interesting and useful contribution from the point of view of system design.

It is also interesting to observe that the numbers of pixels that are required for better classification are different in both databases. The FERET database includes more individuals and fewer images per person than the YALE database, which is a possible reason why more information is required form the images of FERET. The behavior of the classifiers is consistently similar regardless of the image transformation taken into consideration. We have to admit that we expected larger differences in the performance of the classifiers with respect to image resolution. However, it seems that the methods we implemented achieve sound dimensionality reduction even with images of high-resolution. In terms of computational costs, it is definitely beneficial to consider images with relatively low resolution, as presented, to construct feature spaces.

Based on our experimental findings we would recommend manipulating the images prior to constructing the feature spaces. Image transformations such as contrast enhancement and edge detection can lead to better classification. Contrast enhancement and edge detection can compensate for the effects of illumination. From the experimental results we see that variations in lighting deteriorate recognition rates regardless of the classifier under consideration. We found that edge detection generally leads to better recognition. It is possible that edge detection not only eliminates some effects of lighting but also highlights facial features important for classification. Physical features such as eyes, nose, contour of the face, or hairstyle may play a significant role in distinguishing individuals, leading to higher variances between persons and to lower differences among images of the same individual.

Table 13 and Table 14 present the list of parameters that were required to construct the feature spaces with the better performances. In Isomap, the values of K nearest neighbors required to connect all samples of the feature spaces were relatively large. Lets remember that we aim at using the lowest possible value of K as to connect all samples. The facts that the values of K are large indicate that the structure of the data (arrangement of the features) in the given spaces is difficult to reveal. Hence the geodesic distances may not completely portray the true distances between samples. Nonetheless the performance of Isomap is very acceptable from the point of view of classification.

Kernel-PCA is a powerful method for constructing feature spaces. However, its performance depends greatly on the kennel function under consideration. The polynomial kernel of second order shows a general improvement over the other two kernels considered in this study. In fact kernel-PCA with the polynomial kernel delivers some of the lowest error rates.

From all the methods taken into consideration, we see that Fisherfaces, Fisher-Isomap, and kernel-PCA with polynomial kernel deliver better performance in face recognition. On the other hand, kernel-PCA with sigmoid kernel consistently performed the worst in both databases.

The fact that the lowest error rates were computed using non-linear methods for the construction of feature spaces suggest that there is important discriminatory information impossible to extract by linear models. Methods such as Eigenfaces of Fisherfaces perform dimensionality reduction by means of liner correlations among pixels and images. However the entire discriminatory information is not portrayed in the extracted data.

The lowest error rate we computed considering the YALE database is 3.73%, it is provided by kernel-PCA with a polynomial kernel of second order and edge images. The lowest error rate taking into account the FERET database is 10.00%, it is delivered by Fisher-Isomap using contrast-enhanced images.

7.5 Conclusions

This investigation has uncovered and quantified the relationship that exists between classifier's performance vis-à-vis anticipated levels of image qualities and transformations. The experimental evaluations are presented in terms of error rates and standard deviations. Image quality is specified as image resolution. We included two image transformations frequently implemented in face recognition, namely contrast enhancement and edge detection. We based our findings on an extensive suite of experiments carried out on two well-known environments in the area of face recognition, such as the FERET and YALE databases. Various linear and non-linear methods capable of sound dimensionality reduction have been implemented and investigated.

Overall we present experimental evidence suggesting that high-quality images (higher image resolution) do not necessarily lead to better recognition rates. The reported data indicates that a relatively small number of pixels contain most of the discriminatory information necessary for face classification. Our findings have direct implications to the design of face recognition systems, it is possible to reduce computational costs without loosing recognition performance. We have showed some adequate image resolutions for two face recognition scenarios: one including a large number of individuals with few images per person, ant the other depicting a relatively small number of people. We show that image transformations, such as contrast enhancement and edge detection generally lead to better performance by the classifiers. Contrast enhancement led to a reduction in the error rates in Eigenfaces by 7.1% when considering images from YALE, and by 6.17% when considering images from FERET. Edge detection improved classification in Eigenfaces by as much as 18.5% when using FERET, and by 7.87% when using YALE.

Chapter 8

AGGREGATION OF CLASSIFIERS BASED ON IMAGE TRANSFORMATIONS [4]

This section describes a thorough investigation regarding the use of collective knowledge of independent classifiers (experts) in the area of face recognition. We formulate a hypothesis and provide experimental evidence supporting it. This is that different image transformations can offer unique discriminatory information useful for face classification. We quantify the effect of various image transformations in terms of produced error rates. We investigate on the discriminatory capabilities of the resulting classifiers. In particular, we are concerned with fundamental ways of deterioration of visual information (illumination conditions). Two major objectives are delineated as follows:

1. To assess the impact, quantified in terms of classifier's performance, of image transformations in various face classifiers. Both linear and non-linear transformations are considered as vehicles to develop some meaningful feature spaces.
2. To evaluate the quality of collective knowledge, gathered by means of aggregation methods, of feature spaces constructed from various image transformations.

We cover an evaluation of individual and combined classifiers operating in feature spaces constructed from various image transformations. We also elaborate on the conditions at which the classifiers and architectures are most likely to fail or thrive from the point of view of classification.

We envision that the combination of different features (variables) can lead to a better description of individuals, hence revealing relevant distinctive information between persons. Image transformations such as contrast enhancement and edge detection can provide unique discriminatory information when it comes to face classification. Contrast enhancement reduces the impact of lighting and occurrence of shadows in face classification. Edge detection can help emphasize facial characteristics suitable for classification, such as face contour, location and shape of the eyes, nose, mouth, and other relevant visual cues that make a person unique. In this study we consider histogram equalization for contrast enhancement, and Sobel operators for edge detection. Careful combination of information contained in constructed features can lead to the enhanced robustness of the face recognition systems.

Aggregation of classifiers takes place at the decision level by means of majority voting and the Bayesian product rule. Various linear and non-linear methods are implemented for constructing meaningful feature spaces, such as PCA, LDA, Kernel- PCA, and Isomap.

---

[4] A version of this chapter has been submitted for publication.
Jarillo, Pedrycz, Reformat, "Aggregation of Classifiers Based on Image Transformations in Biometric Face Recognition", *Submitted for publication to Machine Vision and Applications on November 2006*, 35 pages.

To our knowledge, this is the first investigation putting together a comprehensive assessment of numerous face classifiers and combined experts in the presence of image transformations; this is, considering both, linear and non-linear methods for constructing feature spaces. We offer evidence of classification improvement over traditional methods as well as design guidelines and recommendations for constructing systems that exhibit robustness against lighting disturbances.

8.1 Previous related work

Over the past years, several approaches involving image transformations, classifiers, and construction of feature spaces have been investigated with respect to classification rates. Particular interest has been given to dealing with lighting conditions and other types of noise. For example, Jarillo, *et al.* [39] presented an extensive investigation on the design of face classifiers capable of operating in the presence of deteriorated (noise affected) visual information. The authors gave attention to Eigenfaces and Fisherfaces methods. They quantified the effects of various noise models on face classifiers. The authors arrived at several design guidelines helpful for constructing classifiers in anticipation of visual disturbances and their intensities. They validated their findings using 200 individuals from the FERET database. Gao, *et al.* [27] implemented the so-called Line Edge Map (LEM). LEM approach performed significantly superior than Eigenfaces in identifying faces with slight appearance variations. LEM approach is also less sensitive to pose variations than Eigenfaces but more sensitive to large facial expression changes. The authors claim that the LEM approach is also more robust to lighting variations than the edge map approach. Yilmaz, *et al.* [105] worked on Eigenhills to overcome problems of illumination. They introduced "hills", which they obtained by covering edges with a "membrane" (mask). Each hill image is then described as a combination of most descriptive eigenvectors, called "Eigenhills", spanning hills space. They reported an improvement of about 7% over the Eigenfaces method. Takács, *et al.* [94] introduced a methodology based on edge images and a modified Hausdorff distance. Their approach operates on edge maps and derives holistic similarity measures using a modified Hausdorff distance, called M2HD. They tested their approach on images of 150 individuals of the FERET database. They report recognition rates of around 92%. Xie and Lam [101] presented the effectiveness of local histogram equalization in face recognition. They tested their approach using images from the YALE database, only those images showing neutral expressions. They report classification rates of about 90% within the framework of Eigenfaces. Belhumeur, *et al.* [10] described "Fisherfaces", a method for constructing feature spaces that shows robustness against lighting conditions. They compared the performance of Fisherfaces against Eigenfaces. They presented evidence of better performance of Fisherfaces to overcome variation in lighting and facial expression.

In face recognition, like in any other pattern recognition problem, it is constructive to combine the strengths of different architectures and/or features towards improving classification rates. The fusion process should include sources of information that complement each other in order to improve the quality of classification. The fusion process may solve the problem of local conflicting decisions and enhance the global accuracy of overall results [83]. In regard to fusion of classifiers, Zhang, *et al.* [108] proposed a method that combines the face image and its Gabor transformation at the feature and matching levels. At the representation level they combined PCA features and LDA features, and at the confidence level, they experimented with the sum and product rules. They concluded that gray level intensities and Gabor features provide different information of identity that can complement each other. Mu, *et al.* [73] explored the performance of summing, voting, and weighted voting methods in combining local distances into the final decision. They proposed a classification method based on weighted voting that allows for a local window to cast a set of weighted votes. They validated their findings using the FERET database. They concluded that weighted voting outperforms simple voting. Khan, *et al.* [43] developed a face recognition method based on PCA and Directional Filter Bank (DFB) responses integrated with a voting algorithm. They used cross correlation as a measure to compare the outputs of various classifiers. The authors tested their approach using the ORL database and reported classification rates of about 96%. They concluded that normalized correlation decision fusion outperforms majority voting. Ivanov, *et al.* [33] explored different strategies for classifier combination (majority voting, sum, and product rules) within the framework of component-based face recognition, they combined information at the feature and classifier levels. They concluded that the product rule outperforms the other two combination strategies.

Other investigations, yet in different areas other than face recognition, where aggregation of classifiers has lead to improved classification rates include the work by Nam, *et al.*[74]. They combined clusters' decisions by aggregation methods. Pham [80] reported on aggregation methods within the context of face detection. Kuncheva [52] presented a comparison between fuzzy and non-fuzzy combination of classifiers. In general, the literature reveals that aggregation of classifiers commonly leads to improved classification performance over single classifiers.

8.2 Aggregation of classifiers

Traditional pattern recognition systems commonly use a single feature space and a classification method to find out the true class of a particular pattern. However finding the true class may be a difficult task in a specified feature space. It has been observed that features and experts of different types could complement each other during classification [47][40]. By combining the opinions of individual experts, a consensus decision (class) can be formed. Various classifier combination schemes have been devised, and it has been experimentally demonstrated that some of them consistently outperform a single best classifier [47]. Ideally, the combination function should take advantage of the strengths of the individual classifiers and compensate for their weaknesses, thus leading to improved classification accuracy [40].

Current literature on aggregation of classifiers reports on two scenarios in which classifiers can be combined. In the first scenario, all experts operate in the same feature space. An example of this category is a set of k-nearest neighbor classifiers, each one using different parameters, such as value of *K* nearest neighbors or distance model. In the second scenario, each classifier operates in independent feature spaces, where the measurements extracted from each pattern are unique to each expert [48]. Many models for combining classifiers take advantage of the intrinsic probabilities of each expert being correct for any given class, assigning weights to each decision accordingly. The foundation for these models is found in the Bayesian decision theory, explained later in this chapter.

This experimental study contemplates the use of features constructed from various image transformations, namely contrast enhancement and edge detection. The aggregation of classifiers takes place at the decision level. Each classifier provides a decision based on its particular expertise, portrayed by the features it takes into account. Figure 31 depicts the general scheme of aggregation of classifiers adopted in this study. There are three image sets created from the same face database Two of these sets are processed by the image transformation operators, creating contrast-enhanced images and edge images. Once the image transformations have taken place, the three image sets are used to construct their corresponding feature spaces, this is by applying the same method of dimensionality reduction to each set. The next step is the classification process, where each classifier casts its decision based on the information portrayed by the given feature space. Finally the proposed class is defined by the aggregation method, which in our case is either majority voting or Bayesian product rule.

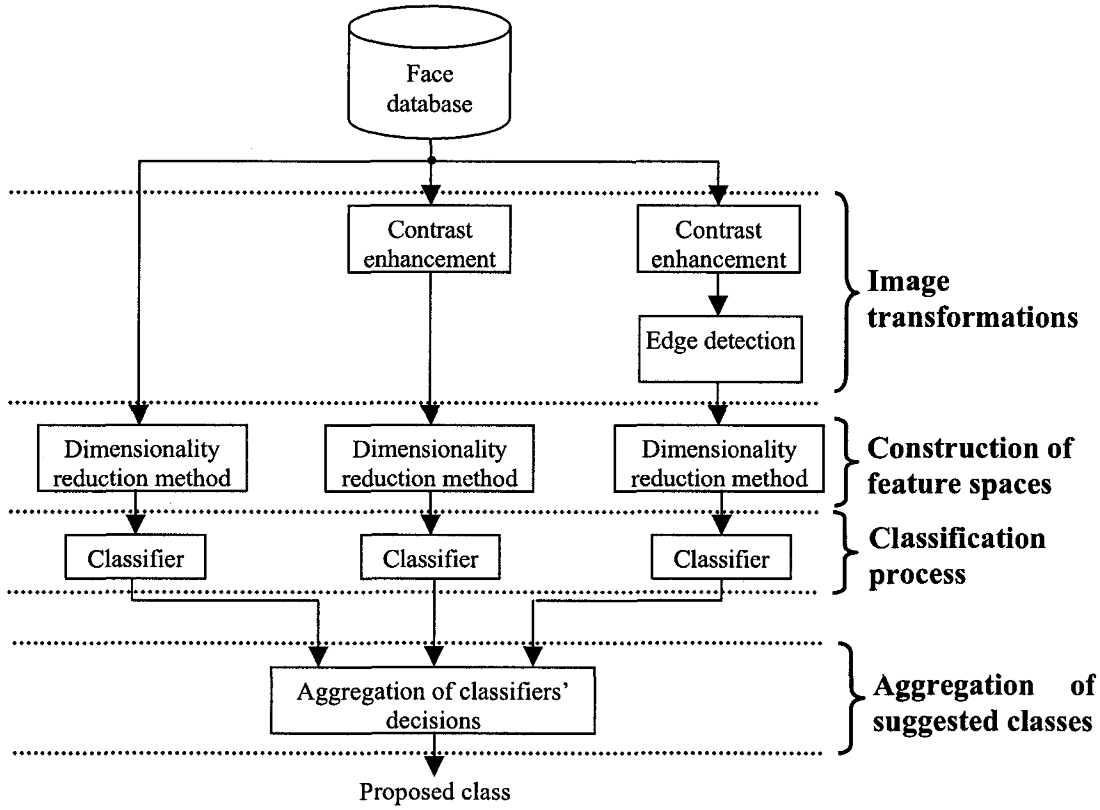


Figure 31 Experimental layout of aggregation of classifiers

8.2.1 Bayesian decision theory – product rule

Bayes decision theory is one of the essential approaches to pattern classification [21]. Let's consider a classification problem where an image **z** is to be labeled as one of C possible classes ($\omega_1, \ldots, \omega_C$). We also consider R classifiers, or in our case feature spaces, each representing the image **z** by a unique feature vector **x**. We denote $\mathbf{x}_r$ as the feature vector being used by the r-th classifier. In other words, pattern **z** is represented in the r-th feature space as $\mathbf{x}_r$. In the feature spaces each class ω_k comes with the underlying class-conditional probability density function $p(\mathbf{x}_r|\omega_k)$. Similarly, let us denote by $P(\omega_k)$ "a priori" probability of occurrence of this class. The Bayes decision rule states that the "a posteriori" probability of feature $\mathbf{x}_r$ belonging to class ω_k is computed as follows

$$P(\omega_k \mid \mathbf{x}_r) = \frac{p(\mathbf{x}_r \mid \omega_k) P(\omega_k)}{p(\mathbf{x}_r)} \tag{16}$$

where,

$$p(\mathbf{x}_r) = \sum_{k=1}^{C} p(\mathbf{x}_r \mid \omega_k) P(\omega_k) \tag{17}$$

The probability density function (pdf) $p(\mathbf{x}_r)$ plays the role of scaling coefficient [16]. This fact makes it possible to concentrate only on the numerator of the above expression Rewriting (16) we obtain

$$P(\omega_k \mid \mathbf{x}_1, \ldots, \mathbf{x}_R) = p(\mathbf{x}_1, \ldots, \mathbf{x}_R \mid \omega_k) P(\omega_k) \tag{18}$$

Therefore, given the features $\mathbf{x}_r$, for $r = 1, \ldots, R$, the pattern **z** should be labeled to class ω_j if the a posteriori probability $P(\omega_k|\mathbf{x}_r)$ is maximum, that is

$$\text{Label } \mathbf{z} \text{ as } \omega_j \textit{ if } P(\omega_j \mid \mathbf{x}_1, \ldots, \mathbf{x}_R) = \max_k P(\omega_k \mid \mathbf{x}_1, \ldots, \mathbf{x}_R) \tag{19}$$

Then, from Bayes decision theory we can formulate ways of calculating all a posteriori probabilities, and therefore suggest the most likely class ω_j for input image **z**. Considering the so-called product rule, we combine the decision of the classifiers as a weighed product of the independent decisions. The weights are estimated according to the belief of each classifier being correct; this is based on their probabilities of correct classification for a given pattern. We estimate the joint probability distribution of the features as

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_R \mid \omega_k) = \prod_{r=1}^{R} p(\mathbf{x}_r \mid \omega_k) \tag{20}$$

by substitution in (18) we arrive at

$$P(\omega_k \mid \mathbf{x}_1, \ldots, \mathbf{x}_R) = P(\omega_k) \prod_{r=1}^{R} p(\mathbf{x}_r \mid \omega_k) \tag{21}$$

leading to the product decision rule

$$P(\omega_j)\prod_{r=1}^{R} P(\mathbf{x}_r \mid \omega_j) = \max_k \left\{ P(\omega_k)\prod_{r=1}^{R} p(\mathbf{x}_r \mid \omega_k) \right\}, \text{ for } k = 1,\ldots,C. \quad (22)$$

In our experiments we computed the probabilities $p(\mathbf{x}_r|\omega_k)$ from the confusion matrix of each classifier. For details see [102]. The confusion matrices are computed using the training and validation sets. The testing set serves to evaluate the aggregation procedure in the testing process; hence the error rates and standard deviations are computed using the testing sets.

8.2.2 Majority voting

The inspiration behind majority voting is based on the assumption that the collective knowledge of a group is superior to those of single experts, provided that the experts have reasonable competence in the field. It is intuitive to say that the quality of the final decision depends on the ability of the experts. In majority voting, each classifier provides a decision (vote) for a given pattern. The class that received the majority of the votes is suggested as the proper class [48]. Since computing weights is not required in this architecture, the votes are computed using only training and testing sets.

8.3 Research environment

Our experimentation takes into account two well-known face databases regarded as benchmarks in the area of face recognition, namely FERET and YALE. Both datasets include variations in lighting conditions and facial expressions. For details refer to section 3.3.

8.4 Experimental setup

The experimental setup is designed to provide evidence regarding performance, in terms of error rates and standard deviations, of independent and combined classifiers. The feature spaces are constructed by means of PCA, LDA, kernel-PCA, and Isomap. Three kernel models are contemplated in kernel-PCA, namely polynomial of second order, Gaussian, and sigmoid with a polynomial kernel of first order. The Gaussian model assumes $\sigma = 70$ for the experiments concerning FERET, and $\sigma = 200$ for those concerning YALE. The parameters of the kernel models were chosen based on our previous investigation described in Chapter 7, the selected parameters provided the lowest error rates from a series of possible values. Isomap was implemented over PCA and LDA feature spaces independently. Hence we have two scenarios for this algorithm that we call Eigen-Isomap, and Fisher-Isomap. The classification task was performed by the nearest neighbor classification rule with Euclidean distance. The aggregation of classifiers contemplates majority voting and Bayesian product rule.

The performance of the classifiers depends directly on the number of variables taken into consideration. In order to find an adequate number of variables, we increased the number of features taken into account by each classifier. The amount that led to the lowest error rate was then selected as proper for that given classifier. In this study, the reported error rates are the lowest we computed considering a particular space and number of variables. For a complete description of the evaluation approach please refer to section 3.1.

The experimental setup comprises a collection of experiments using the FERET and YALE databases. The FERET images were divided into 200 images for training, 200 for validation, and 200 for testing with no overlap (one image per class in each set). A 3-fold cross validation is taken into account to evaluate those architectures using the FERET database. The YALE database contains 11 images per person for a total of 15 individuals. We created 10 random splits with overlap, each having a training, validation, and testing sets of 4, 3, and 4 images per person respectively. The evaluation of the architectures using the YALE datasets is reported as the average and standard deviation computed over the 10 splits. FERET and YALE databases were treated independently throughout the experimentation. For a complete description of the databases please refer to section 3.3.

There are two image transformations implemented in this study: contrast by histogram equalization and edge detection by the Sobel operator. To improve the performance of the edge detection process, the images were pre-processed by contrast enhancement. Both image transformations were applied to both databases, giving place to three major sets for each database. All three major sets from FERET and YALE were subjects to the dimensionality reduction by means of Eigenfaces, Fisherfaces, Isomap, and kernel-PCA. The aggregation of classifiers takes into account those features spaces computed by the same dimensionality reduction method but different image transformations, i.e. no transformation, contrast, and edge detection (see Figure 31). Therefore there is one aggregation algorithm for each dimensionality reduction method. However, it considers three classifiers to come up with the final decision.

Aggregation by majority voting and Bayesian product rule were considered in this study. When it comes to the Bayesian product rule, we are required to compute a set of weights, one for each expert, and to select a proper number of features from the space. The weights were computed using the images in training and validation sets, the number of features were set according to the lowest error rate over the validation sets. The overall performance of the aggregation of classifiers was evaluated over the testing sets. When it comes to majority voting no weights are required, hence the validation sets are not needed. The error rates from each independent expert are compared to the error rates of the combined classifiers. To validate the significance of the results, a statistical t-test [72] is performed on the classification results given by a single expert and by the combination of them.

The experimental setup allows us to evaluate the collective knowledge of the experts. We expect to offer evidence to support the hypothesis that different image transformations provide different discriminatory information useful for classification. And that such discriminatory information can be combined to improve performance over single classifiers.

8.5 Experimental results

There are three feature spaces for each method of dimensionality reduction. The feature spaces emerged from regular images, histogram-equalized images, and edge images. Table

15 shows the number of variables for which the error rate is lowest in each given feature space, this is considering training and validation sets. The data presented in Table 15 comprise YALE and FERET databases.

Table 15 Number of variables that provide the lowest error rates in each feature space

Method	Number of Features					
	No transformation		Histogram equalization		Edge detection	
	YALE	FERET	YALE	FERET	YALE	FERET
PCA	58	124	28	185	57	199
LDA	14	187	14	193	14	196
Eigen-Isomap	47	126	60	185	54	198
Fisher-Isomap	53	191	52	190	60	198
Kernel-PCA (polynomial)	51	170	51	173	37	147
Kernel-PCA (Gaussian)	22	134	23	140	16	119
Kernel-PCA (Sigmoid)	53	166	55	158	60	148

Figure 32 and Figure 33 show the performance, in terms of average error rates and standard deviations, of classifiers using the YALE and FERET database respectively. The data is organized according to feature spaces specified in the horizontal axis. The bars indicate the error rates of single and combined classifiers computed over different feature spaces. Thin lines overlapped with the error rates show the standard deviations. The presented error rates were computed over the testing sets alone.

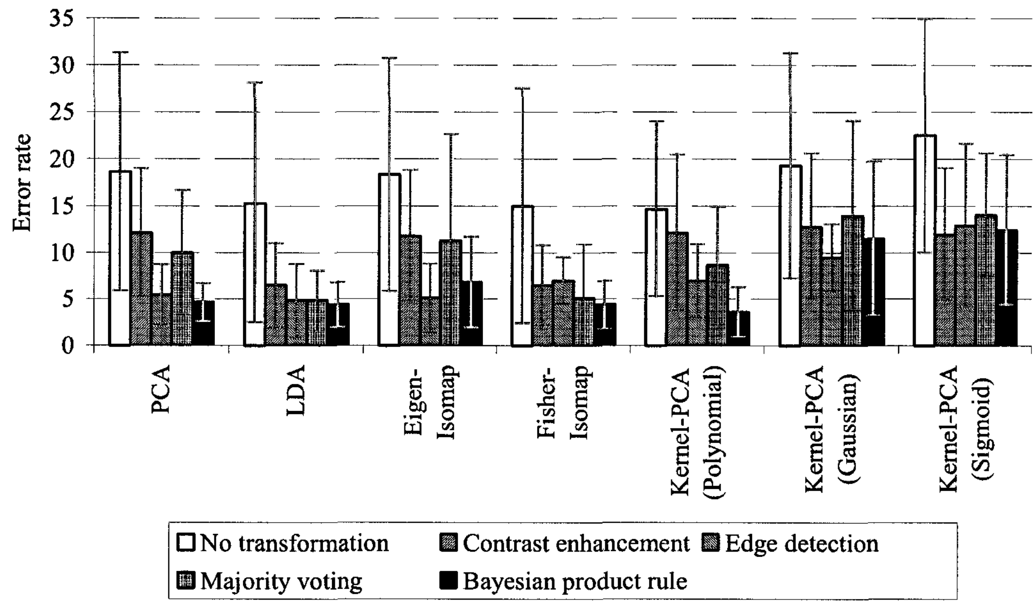


Figure 32 Performance of single and combined classifiers in YALE

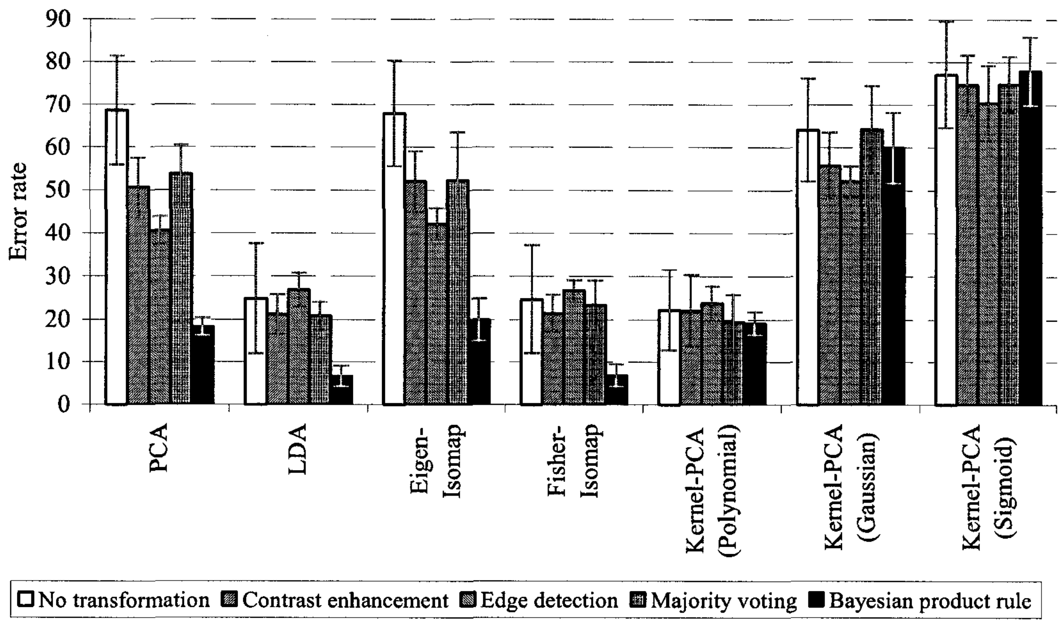


Figure 33 Performance of single and combined classifiers in FERET

To validate our findings, we assessed the statistical significance of differences between reported average error rates using t-test. The objective is to evaluate, from the statistical point of view, whether in fact combined classifiers improve classification rates over the results produced by single classifiers. At this point we compare the error rates of single classifiers using unaltered images (images not affected by any transformation operator), against the outcomes of combined classifiers (majority voting and Bayesian product rule). Table 16 presents the feature spaces and aggregation methods for which we can support the hypothesis that the classification improvement is statistically significant with over 90% confidence level.

Table 16 Feature spaces and aggregation methods showing improved performance

	Majority voting	**Bayesian product rule**
FERET	NONE	PCA LDA Eigen-Isomap Fisher-Isomap
YALE	PCA LDA Fisher-Isomap Kernel-PCA (polynomial) Kernel-PCA (sigmoid)	PCA LDA Eigen-Isomap Fisher-Isomap Kernel-PCA (polynomial) Kernel-PCA (sigmoid)

The experimental results presented in Figure 32 and Figure 33 suggest that image transformations provide relevant discriminatory information in face classification. We observe reduction of error rates when classifying features emerging from histogram equalization and edge detection over features from conventional images. We expected histogram equalization to balance lighting, hence reducing error rates for methods that are susceptible to this type of noise. In Eigenfaces we can clearly see the advantage of histogram equalization, the error rates drop significantly when the images are pre-processed by contrast enhancement, as much as 18% in the case of FERET.

From a different perspective, edge detection is intended to stress out the shapes of facial features, making variances much larger for different individuals, and smaller for images of the same individual. Experimental results suggest that edge detection indeed exposes more relevant discriminatory information.

Our experimental findings reveal that image transformations influence the manner in which information is allocated along each axis of the feature spaces, especially edge detection. Figure 34 portrays an example of the cumulative variances along each axis of the PCA feature space. The feature spaces were constructed from images with and without transformation operators. In this example the PCA feature space was computed using the YALE database. In Figure 34 we can appreciate that the information is allocated along each axis of the feature space in a particular way. For example, features from images with no transformation and from histogram equalization contain most of the variances in fewer variables than features from edge images. The implications of this phenomenon are far reaching.

The influence of each variable in the similarity measurements, portrayed by a particular distance model, can modify the outcome of the classifiers. This issue is particularly important in the design of classifiers susceptible to distance models, such as the nearest neighbor classification rule. An adequate distance model may be obtained experimentally, for instance one could try various models emerging from the Minkowski family of distances [16], which includes the widely used Euclidean distance. The fact that the variances are more evenly distributed over the entire space is reflected in the number of features that provide the lowest error rates, see Table 15.

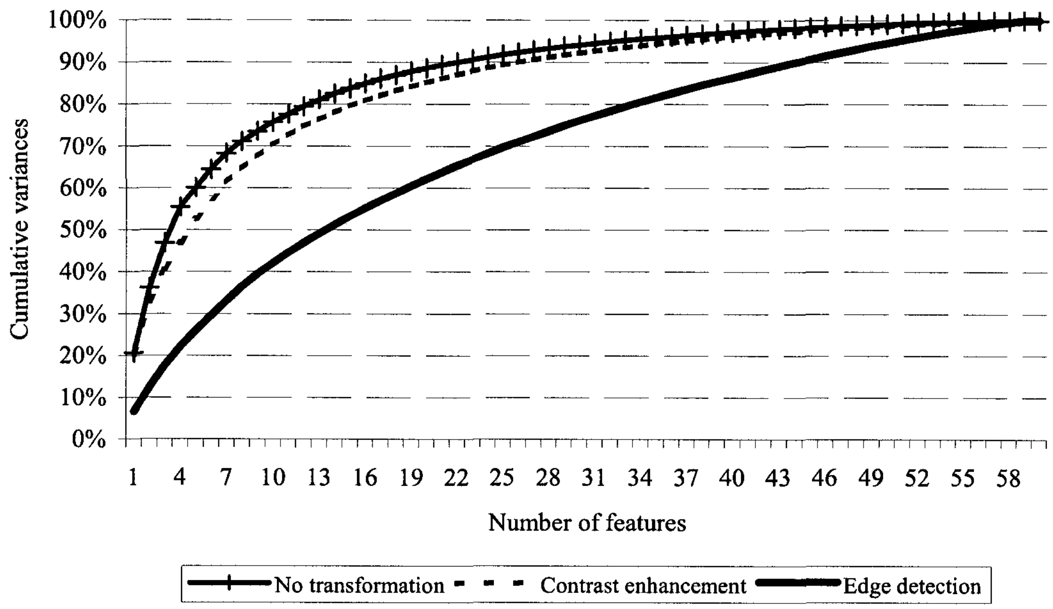


Figure 34 Cumulative variance along each coordinate of PCA feature space

In general, kernel-PCA with a polynomial kernel of second order can lead to low error rates. However it is not helpful with either Gaussian or sigmoid kernels, particularly with large databases such as FERET. The experimental results also indicate that combining classifiers' decisions can significantly improve classification. In general, combining classifiers' decisions by Bayesian product rule leads to lower error rates. Nonetheless majority voting shows some minor improvements in some cases. It is possible that one of the classifiers outperformed the remaining two in several occasions, leading to a more robust classification with Bayesian product rule.

Overall, the lowest average error rate in the YALE database is 3.66%. It is offered by Bayesian product rule and kernel-PCA with polynomial kernel of second order. The lowest error rate in FERET is 6.66% using Bayesian product rule and Fisher-Isomap. Figure 35 shows samples of misclassified images in YALE and FERET databases. The samples come from the aggregation of feature that showed best performances, i.e. Kernel-PCA with YALE and Fisher-Isomap with FERET. It is interesting to see that even though contrast enhancement and edge detection intend to deal with lighting conditions. Some images with such type of noise were still misclassified.

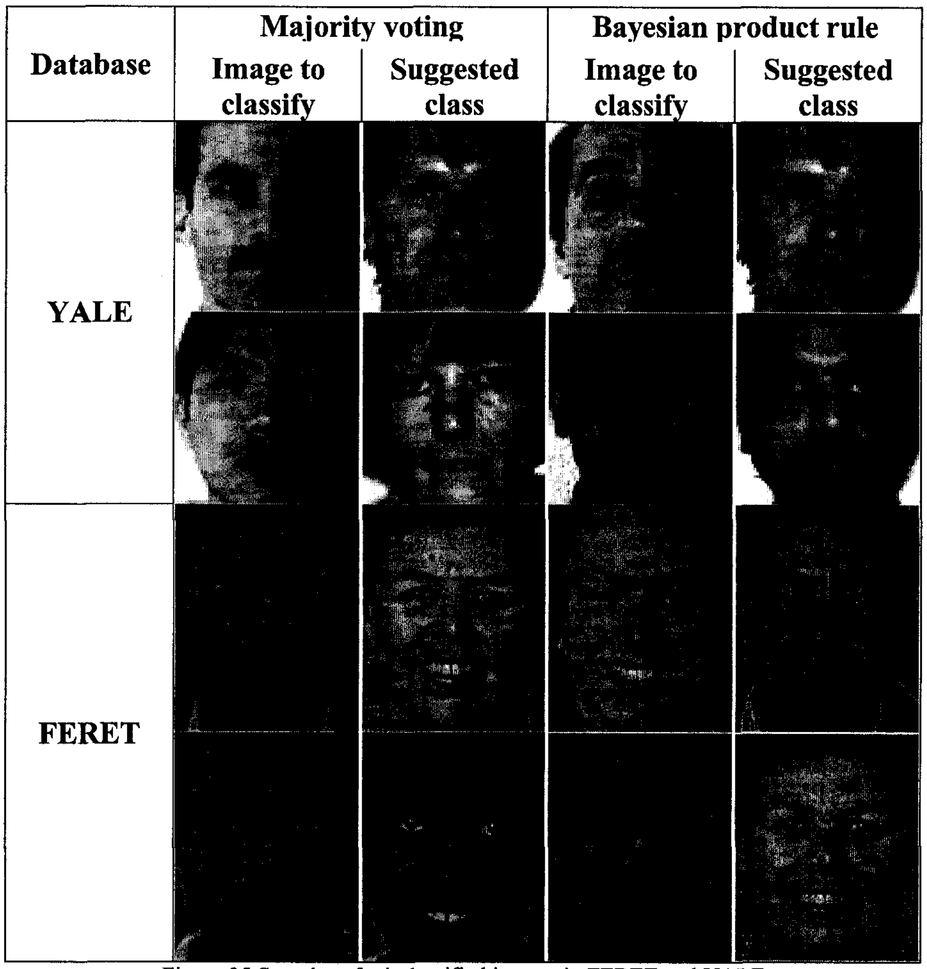


Figure 35 Samples of misclassified images in FERET and YALE

An examination of the errors made by each classifier reveals particular cases where the classification by aggregation can enhance classification performance, and cases where no improvement is possible. Any classifier can produce wrong class assignments for particular images. However, if the other classifiers make the correct classification then the final proposed class could be correct. In other words, the aggregation method combines the strengths of some classifiers to overcome the weaknesses of others. In Figure 36 we present a graphical representation of the unique and mutual errors made by each classifier. The figure was constructed taking into consideration one of the data splits from the YALE database; it considered feature spaces constricted by kernel-PCA with polynomial kernel model (kernel-PCA showed best performance in this database). Each ring shows the number of errors made by one classifier. The overlapped areas depict the number of cases where the errors are mutual. The area in the middle portrays the number of mutual errors among all classifiers. In Figure 36 we see that the classifier that considers no image transformation made 7 unique errors (errors not shared by any other classifier). The classifier that considers contrast enhancement made 6 unique errors, and the classifier that considers features from edge detection made 4. There are 5 mutual errors shared by the classifier that considered images with contrast enhancement, and the classifier that considered images with no transformation. Overall, there are 4 out of 8 cases where all classifiers made exactly the same errors, they proposed the same wrong classes for 4 identical images of the test set. We can see now that these 4 cases will be misclassified regardless of the aggregation technique we implement. Nonetheless there can be improvement over the remaining 4 images if the aggregation method is carefully designed.

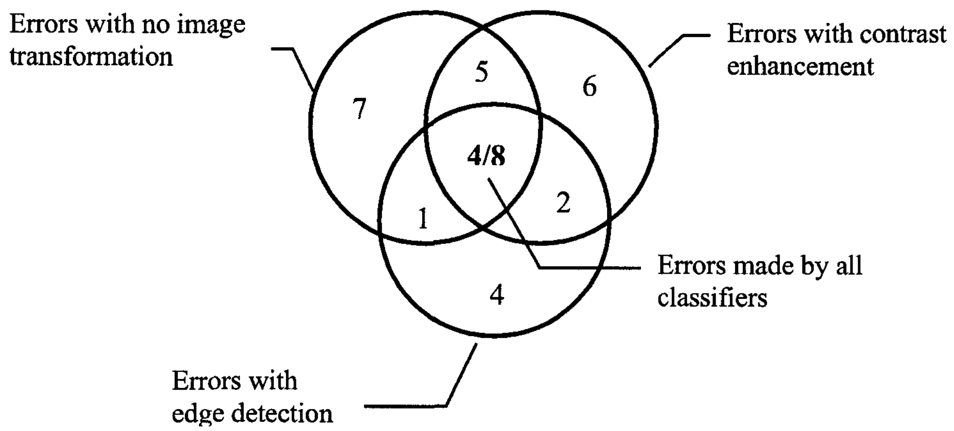


Figure 36 Graphical representation of unique and mutual errors made by independent classifiers (kernel-PCA with polynomial kernel) using YALE

Figure 37 depicts the number of unique and mutual errors made by each classifier. The figure was constructed taking into consideration one of the validation splits from the FERET database; it considered feature spaces constricted by Fisher-Isomap (Fisher-Isomap showed the best performance in this database). For this particular scenario, there are no shared errors among all classifiers (innermost area), therefore there is the possibility of achieving correct classification for all cases if an aggregation method is carefully constructed. At this point it is worth mentioning that Figure 37 shows the errors made over one validation split only, hence the overall average error rate may still be higher than zero.

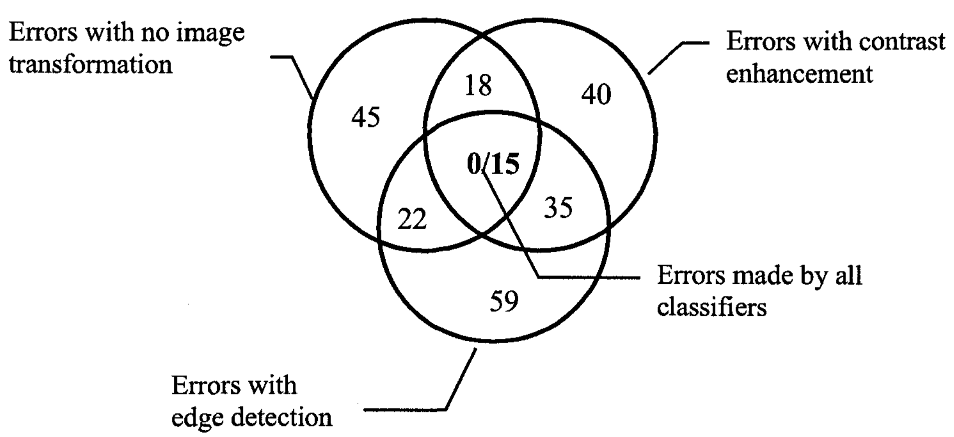


Figure 37 Graphical representation of unique and mutual errors made by independent classifiers (Fisher-Isomap) using FERET

8.6 Conclusions

We have presented compelling experimental evidence that supports the hypothesis that different image transformations (such as contrast enhancement and edge detection) provide additional information useful for classification. Similarly, this discriminatory information can be of relevance when carrying out aggregation of classifiers. Various methods for constructing meaningful feature spaces have been considered, namely PCA, LDA, Kernel-PCA, and Isomap. Majority voting and the Bayesian product rule are the two constructs that realize a combination of experts' decisions. We presented cases where improvement can be achieved by a careful design of aggregation methods. One should stress that there are also situations where no improvement has been reported. The experimental scheme involves two well-known face databases regarded as benchmarks in the area of facial recognition, namely YALE and FERET.

In general, combining features by means of the Bayesian product rule leads to the improved performance of the resulting classifiers. The lowest error rates on YALE were provided by kernel-PCA with polynomial kernel of second order, and by Fisher-Isomap on FERET. The implications of image transformation in face classifiers can significantly affect classification performance. We found out that features constructed form edge images tend to depict an even distribution of discriminatory information over the entire space. Careful design of classifiers should be taken into consideration when dealing with image transformations, particularly in the case of choosing similarity measurements (distance models).

Chapter 9

IMPORTANCE OF FEATURES IN FISHER SPACE – AN EVOLUTIONARY APPROACH

Most approaches to face recognition involve some sort of transformation of the original face images (inputs representing pixels values) to a meaningful feature space. Given a collection of projected images (feature vectors) we proceed to the identification process by means of some classifier(s), typical approaches adopt a nearest neighbor rule. During classification, the variables forming the feature vectors assume equal "importance", this is the similarity measurement (distance model) does not distinguish between variables with various discriminatory powers. It is intuitive to think that classification performance would improve if the classifier took into consideration the discriminatory capacity of the variables. To this end, we have developed a face recognition system that reveals the discriminatory powers of the variables and that uses them in order to improve classification. In addition, our system proposes an adequate similarity measurement to distinguish between individuals in the given feature space.

This chapter presents an extensive experimental investigation on an evolution-driven approach to face recognition. This investigation departs from a well-established method for constructing meaningful feature spaces, namely Fisherfaces. Its intrinsic characteristic of separating the classes and its capability of sound dimensionality reduction make it an ideal platform to build upon. Evolutionary optimization is carried out by Genetic Algorithms (GA). A GA ranks the variables of the feature vectors according to their "importance" for classification, and at the same time, produces a suitable distance model capable of separating the classes (individuals). The distance models are derived from the Minkowski family of distances, described in [16] and in section 3.2.2. Evolution is driven towards reducing error rates.

Throughout the chapter we present experimental evidence supporting the usefulness of this approach. We comment on advantages and drawbacks of this method, particularly in terms of computational costs and error rates. We provide design recommendation regarding suitable environments and feature spaces for this technology.

9.1 Previous related work

Various attempts to improving performance of face classifiers have been made with the help of Genetic Algorithms. In this section we provide some representative examples the architectures proposed in the past. In 1998, Liu, *et al.* [59] worked on an approach called Optimal Projection Axes (OPA) for face recognition. OPA works by searching through all the rotations defined over PCA subspaces. The authors expected better performance from non-orthogonal bases over orthogonal ones. Evolution is driven by a fitness function defined in terms of performance accuracy and class separation (scatter index). Accuracy indicates the extent to which learning has been successful, while the scatter index provides an indication of the expected fitness on future trials. The authors tested their approach over 1107 images corresponding to 369 subjects from the FERET database. Liu, *et al.* showed that OPA delivers improved performance over Eigenfaces. The authors report a top performance of 92.14% recognition.

In 2002, Yankun, *et al.* [104] described an approach to face recognition based on kernel-PCA and Genetic Algorithms. The authors included a polynomial kernel of various degrees for their experimentation. For the classification task they employed linear Support Vector Machines (SVM). The authors tested their approach over their own database consisting of 1400 images of 70 individuals, and over the ORL database. The task of the GA was to select suitable variables for classification. The chromosomes consisted of binary genes, where a value of 1 indicated inclusion of a variable and a value of zero indicated exclusion. The authors report a recognition rate of about 98% on their own database with a polynomial kernel of degree 4, and an error rate of 1.63% over the ORL database. For the latter they took into consideration a polynomial kernel of degree 5.

In 2004, Xu, *et al.* [103] presented a face recognition system capable of selecting suitable features for classification. With the aid of GA they were able to extract relevant information from features derived from Independent Component Analysis (ICA). The chromosomes comprised binary genes that specified which variables to extract. Classification was performed by means of the nearest neighbor classifier and Cosine similarity measure. The authors tested their method using images of the YALE database. Xu, *et al.* delivered a comparison between the performance of Eigenfaces, FastICA, M-FastICA, and FastICA with GA. Their results show better performance of FastICA with GA, about 90.32%.

In 2005, Zheng, *et al.* [109] introduced an approach involving GA as a vehicle to identifying the variables suitable for classification. They considered PCA and then LDA for their approach. As initial step, the authors developed the so-called GA-PCA, which extracts useful features from the PCA space, then they utilized GA once more over the variables produced by LDA. LDA took into consideration the variables previously extracted by GA from PCA space. The authors compared the performance of their approach to that of Fisherfaces. The Face Recognition Technology (FERET) and Carnegie Mellon University Pose, Illumination, and Expression (CMU PIE) databases are used for evaluation. Experimental results showed an improvement of about 5% over Fisherfaces.

In 2006, Liu, *et al.* [63] presented their work on face recognition using Genetic Algorithms. They described a nonlinear Evolutionary Weighted Principal Component Analysis (EWPCA) based on Genetic Algorithms. Similar to LDA, the EWPCA maximizes the ratio of between-class variations to that of within-class variations, and achieves better classification performance than that of traditional PCA. The authors entrusted Genetic Algorithms to select optimal weights for the EWPCA. In face recognition, Evolutionary facial feature obtained by performing EWPCA is used as the representation of original face images. Liu, *et al.* evaluated their proposed algorithm on the Cambridge ORL face database and a combo database consisting of ORL, Yale, and UMIST databases. They showed that

EWPCA could outperform PCA, kernel PCA and LDA. The authors reported an error rate of about 95% over the ORL database, and about 92% over the combo database.

9.2 Genetic algorithms – An optimization tool

This section introduces Genetic Algorithms as an optimization tool. It provides the required background before elaborating on the proposed face recognition architecture. Many Evolutionary algorithms have become popular tools for searching, optimization, machine learning, and for solving problems. These algorithms mimic evolution in order to discover solutions to complex problems [100].

Genetic algorithms were developed in the US by John Holland and his students in the 1960s, 1970s, and 1980s [100]. An interesting line to describe such algorithms reads: "Computer programs that "evolve" in ways that resemble natural selection can solve complex problems even their creators do not fully understand" [122]. As of today GA remains the most recognized form of evolutionary algorithms [100].

Typically an optimization application requires finding a set $\mathbf{x} = \{x_1,..,x_n\} \in M$ of free parameters of the system under consideration, such that certain quality criterion, usually called the objective function, is maximized, or equivalently minimized. This is $f(\mathbf{x}) \rightarrow \max$ [7]. The objective function might be given by real-world systems of arbitrary complexity or by an analytical expression. In general, a solution to the global optimization problem stated above requires finding a vector $\mathbf{x}$ such that $\forall \mathbf{x} \in M : f(\mathbf{x}) \leq f(\mathbf{x}^*) = f^*$ [7].

In genetic algorithms, the potential or possible solutions for a specific problem form what is called a "population", which is simply a collection of "individuals", also known as "chromosomes" or "genotypes". The artificial evolution mimics natural selection in the context of probabilistic operators, namely mutation, selection, and recombination. These operators seek for the evolution of the individuals (potential solutions) towards better fitness values - survival of the strongest. The fitness value is a direct indication of the objective function's value of an individual in relation to a fitness function to be optimized.

In order to represent the potential solutions (chromosomes) of the original problem at hand, one has to encode (transform) the original problem into a format susceptible to genetic computations. In the opposite direction, an inverse transformation is accomplished by a decoding mechanism that permits moving from the GA space into the original search space. Generally speaking, the encoding/decoding mechanism comprehends three possible scenarios as follows [16]:

- *1-to-1* mapping
- *n-to-1* mapping
- *1-to-n* mapping

Such mechanism is well illustrated in Figure 38 (illustration from [16]).

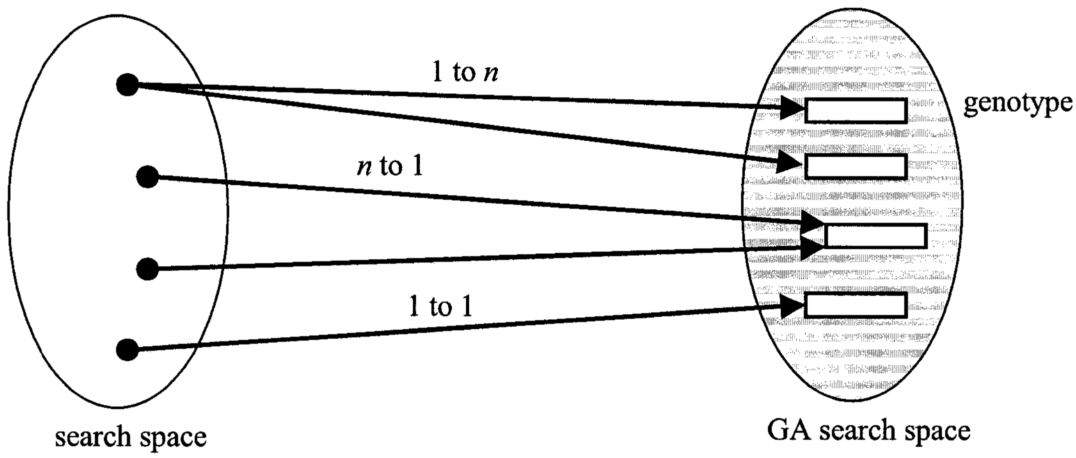


Figure 38 GA encoding – decoding scheme

The GA search space is composed of a string of symbols. In their simplest form they can comprehend an alphabet of 2 elements {0, 1}. However, the alphabet can be expanded as necessary to the point of using real numbers (floating point encoding). The design of the chromosomes is fundamentally linked to the nature of the problem, and it should represent a solution adequately and completely.

The mutation operator introduces innovation into the population by generating variations of individuals [7]. It changes the value of an element of the chromosome, for instance, if the chromosome is formed by binary variables, then the operator changes the value of the variable to its complement. At this point, the mutation rate denotes the probability at which the individual bits become affected. For instance, a mutation rate of 5% when applied to a population of 500 chromosomes, each being of 20 variables, means 5% of 1,000 variables being changed [16]. If the GA search space is defined using a more extended set of symbols (such as floating point encoding), the mutation operator can modify the chromosomes' variables content by some random increments [16]. Another approach that may be useful for covering more drastically the search space is to change the value of one or few chromosomes' variables by a totally new random value.

The recombination operator, usually called "crossover", performs an information exchange between different individuals in the population [7], hopefully inheriting relevant part of an adequate solution to the offspring. A one-point crossover identifies two chromosomes from the population and randomly selects a position in the chromosomes at which they interchange their content. Continuing with examples of chromosomes of binary variables, we present Figure 39 to depict the process.

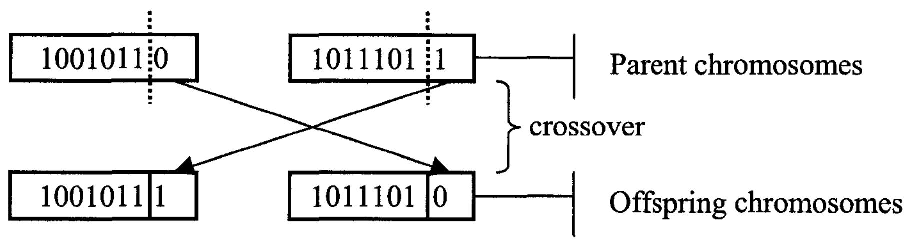


Figure 39 One-point crossover operator

The direct effect of the crossover operator in the evolutionary process is the diversification of individuals that form the new population. The crossover is denoted in terms of probability at which the elements of the chromosomes are affected. The higher the probability, the more individuals are affected by this operator [16]. The crossover is regarded as a fundamental mechanism of a GA [16]. Within the same context of crossover, and considering real numbers as the plausible alphabet for the chromosomes' variables (floating point encoding), we present the following mechanisms to do the task [16]:

- *Linear recombination.* It considers two parent chromosomes **x** and **y**, similarly there are two offspring **x'** and **y'**. The recombination occurs as follows

$$\mathbf{x}' = A\,\mathbf{x} + (1 - A)\mathbf{y} \tag{21}$$
$$\mathbf{y}' = A\,\mathbf{y} + (1 - A)\mathbf{x} \tag{22}$$

  where A is constant assuming values within the range [0-1].

- *Flat crossover.* It is defined as

$$z_i = \mathrm{U}(\min(x_i, y_i), \max(x_i, y_i)) \tag{23}$$

  where $i = 1,\ldots,M$. where M is the length of the chromosome vector. U(a, b) is a random variable with a uniform distribution function that is defined over [a, b], and **z** is the resultant offspring.

- *Simple crossover.* It follows a similar idea as the crossover defined previously. For Boolean variables it is

$$\mathbf{z} = [x_1, x_2,\ldots, x_i, y_{i+1},\ldots,y_M] \tag{24}$$

  where i is randomly chosen within the range of the chromosome [1-M].

- *BLX-a crossover.* It is defined as

$$z_i = \mathrm{U}(\min(x_i, y_i) - \mathrm{I}_i a, \max(x_i, y_i) + \mathrm{I}_i a) \tag{25}$$

  where $\mathrm{I}_i = \max(x_i, y_i) - \min(x_i, y_i)$, and a assuming values between [0-1].

The evolutionary process measures the "quality" of the population. No doubt it is crucial for the proper evolution of the solutions. At this point the fitness function evaluates the chromosomes independently and assigns a corresponding fitness value (strength) to each of them. An adequate fitness function will allow the proper selection of the best solutions to the problem at hand. If one is interested in the general "strength" of a population, an average fitness value can be computed. A common procedure is to plot the average strength of the population according to the number of generations (iterations), which depicts the evolution of the solutions.

The selection operator is the driving force behind the process of evolution, by preferring better, or most fit, individuals to survive and reproduce after the members of the next generation are selected [7]. The selection operator mimics the process of natural selection (survival of the strongest individuals) in a population. A standard mechanism is the roulette wheel selection described as follows:

All fitness values of the individuals forming the population are normalized to a maximum value of 1. Then the normalized values can be regarded as probabilities in the following manner [16]

$$prob_i = \frac{fitness_i}{\sum_{j=1}^{N} fitness_j} \tag{26}$$

where the sum of the values in the denominator characterize the total fitness of population P. We then form a roulette wheel whose sectors are formed to reflect the probabilities of the chromosomes in the population, as illustrated in Figure 40. If we spin the wheel N times and select chromosomes from it, we arrive to a new selected population. Naturally, the chromosomes with higher fitness values (higher probabilities of being chosen) will be selected more often than those with lower fitness values.

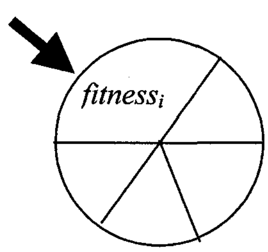


Figure 40 Roulette selection

To guarantee evolution, one can also implement an "elitist" mechanism, which extracts the strongest, or few strongest, chromosomes from the present population and directly copies them into the new population. The process is carried out without any modification to the selected chromosomes whatsoever.

The roulette selection works well in many cases. However, its weakness is present in complex fitness functions. If the best chromosome takes over, then the entire population may be confined to a specific area, therefore finding a local optimal solution. To alleviate the drawback of the roulette selection, one can apply the scaling and ranking mechanisms, which helps keep competent individuals and at the same time maintain diversity in the population. In general, the scaling operator transforms the fitness value denoted by *fit*, into a new expression *fit'*, where *fit'* = *f*(*fit*) and *f* being the scaling transformation. Some transformations are outlined next [16]:

- *Linear scaling.* It transforms the value of fit by a linear function as follows

$$fit' = a \cdot fit + b \tag{27}$$

  where a and b are selected so that average chromosome receives one offspring and best receives certain number of copies.

- *Dynamic linear scaling.* It changes the parameters of the transformation over the course of evolution. This is expressed as

$$fit' = a \cdot fit + b(t) \tag{28}$$

  where $b(t)$ varies with the present population. Particularly one can assume $b(t)$ as the negative value of the fitness of the weakest individual.

- *Sigma truncation.* It is an approach that modifies the values of the fitness function according to the characteristics of the entire population, this is

$$fit' = fit - (\overline{fit} - c\sigma) \tag{29}$$

  where $\overline{fit}$ and σ are the mean value of the fitness function across the entire population, and the standard deviation respectively.

- *Power law scaling.* It transforms the original fitness values of the chromosomes according to

$$fit' = fit^{\alpha}, \text{ or similarly } fit' = (a \cdot fit + b)^{\alpha} \tag{30}$$

  If α tends to zero, it produces a mechanism of random selection. On the opposite case, more attention is focused on the individuals with higher fitness values, and tends to play a more essential role in the construction of the next population.

The algorithm of the complete process can be expressed in a loop that mimics evolution as it roughly occurs in nature by implementing the operators of recombination, mutation, fitness evaluation, and selection. The process iterates through the loop for a specific number of times (generations), or until a proper solution is found.

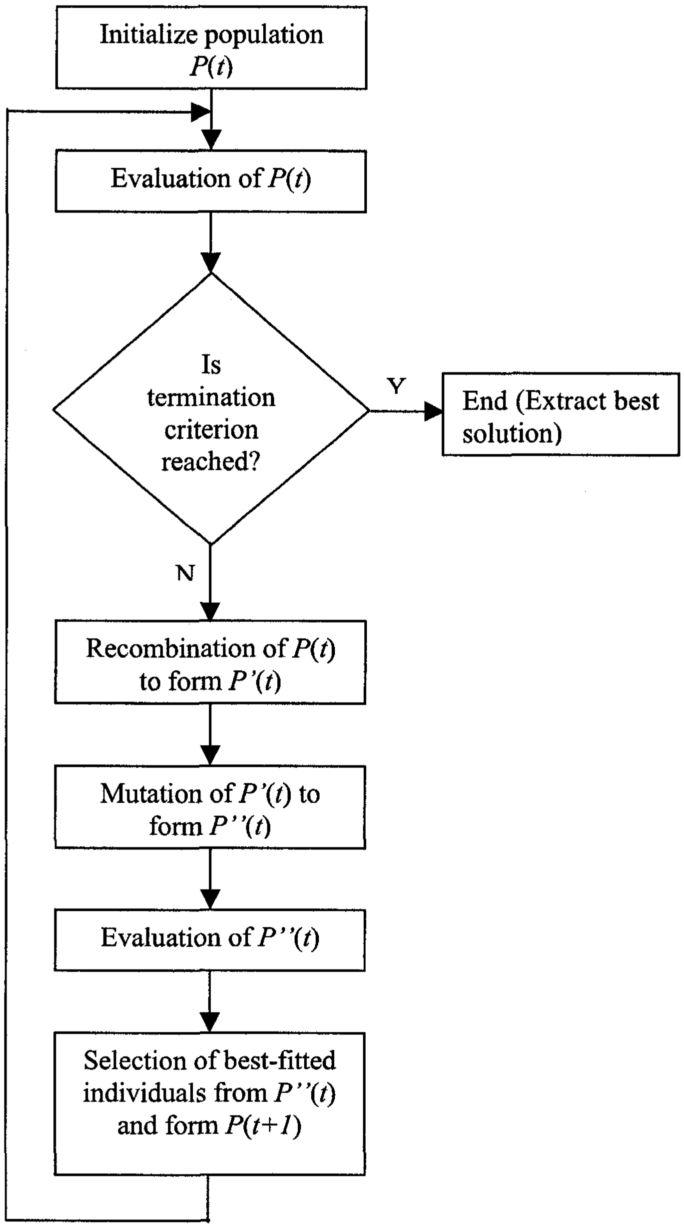


Figure 41 Genetic Algorithm scheme

Figure 41 depicts the general scheme of a GA. $P(t)$ denotes a population of potential solutions in generation t. $P''(t)$ stand for an offspring population created after recombination and mutation of population $P(t)$. The offspring individuals are evaluated by calculating the objective function values $f(\mathbf{x})$ for each potential solution $\mathbf{x}$ that is present in population $P''(t)$. Finally the selection is performed based upon the fitness value of every individual to evolve towards a better solution. The GA process starts by randomly initializing the chromosomes of the population $P(0)$.

The evaluation of all individuals takes place by means of the fitness function, if a satisfactory criterion is met then stop and extract the best solution (the one with highest fitness value if the criterion is to be maximized) from the pool. If no satisfactory solution is found, then continue the recombination operator (crossover) to generate a population $P'(0)$ with the consequent offspring.

The mutation process is used to form a population $P''(0)$, evaluate every chromosome in $P''(0)$ and carry on with the selection mechanism. The newly formed population $P(1)$ is then evaluated to check if a proper solution has been found, if not then continue with the iterative process until the termination criterion is satisfied.

9.3 Research environment

The research environment considers images from the YALE database. We count 11 images for 15 individuals with variations of gesture and illumination. All classes are taken into consideration during training and classification. Classification performance is expressed in terms of error rates via the nearest neighbor classification rule. Details of the research environment can be found in section 3.1.

9.4 Experimental setup

The experimental setup is designed to reveal the "importance", from the classification point of view, of the variables forming the feature vectors in Fisherfaces space. It also intends to deliver a suitable similarity model for face classification. The distance models are derived from the Minkowski family of distances, where the form of the distance function is specified by a parameter p. The Minkowski family of distances is presented in (1) in section 3.2.2.

The importance of the variables forming the feature vectors is established by a weight constant that can assume any real value between zero and one, the higher the weight the higher the importance of the corresponding variable. The nearest neighbor classifier then considers a weighted distance measurement taking into account the proposed weights and distance model. A genetic algorithm is entrusted the task of discovering a suitable distance model (defined by an integer value of parameter p) and the weight values (importance) of the variables of the feature vectors. Evolution in the GA is driven toward reducing error rates. The fitness function to maximize takes the form

$$fitness = \frac{1}{1 + error\_rate(p, \mathbf{w})} \tag{31}$$

were the error rate is a function of the Minkowski distance defined by p, and the set of weights $\mathbf{w}=\{w_1,\ldots,w_{C-1}\}$ associated to the variables of the feature vectors. Note that we can have at most C-1 weights in the Fisherfaces space. Parameter p is allowed to assume any integer value between 1 and 10.

In our investigations we contemplate 10 random splits of the YALE database, the performance is specified as error rates and corresponding standard deviations. We consider two scenarios for our evaluations:

- *Scenario 1.* We include 6 and 4 images for training and testing sets respectively. The GA finds the weights and the parameter p taking into consideration both, training and testing sets. Error rate is computed over testing sets. The performance is presented as average error rates and standard deviations over the 10 random splits.
- *Scenario 2.* We include 4, 4, and 3 images for training, validation, and testing sets respectively. The weights and the parameter p are found taking into account the training and validation sets. The evaluation is performed over the testing sets. For comparison purposes, we also present the average error rate and standard deviation over the validation sets.

Any experimentation involving a GA requires a number or repetitions for each experiment. This is to avoid biased results due to the randomness involved throughout process. In our experiments we considered 10 repetitions for the 10 splits of the YALE database, hence resulting in a suite of 100 experiments for each scenario. The proposed weights are expected to be somewhat different after every repetition of the experiments within a narrowed range. In scenario 2, the evaluation over the test sets takes into consideration the average of the weights proposed by the GA over the validation sets. As for the distance model (value of parameter p) we took into consideration the value that was suggested most of the times throughout the 10 repetitions over the validation sets.

The structure of the chromosome involves a vector of weights and the parameter p as illustrated in Figure 42. They are treated independently through the crossover and mutation operations. We considered linear recombination as crossover operator and a random mutation. In order to compute a valid value of p, the outcomes of the crossover and mutation operators were rounded off to the closest integer.

w_1	...	w_{C-1}	p

Figure 42 Chromosome representation

The GA followed the "roulette selection" mechanism in order to select the chromosomes for the next generation. The probabilities of crossover and mutation were set to 70% and 7% respectively, such probabilities led to a somewhat faster evolution of the population. The GA comprised a population of 500 chromosomes. The evolution ran for 1000 generations or until the fitness value of a chromosome reached 1.

9.5 Experimental results

In what follows we expand on the experimental results of both scenarios. We present the average weights for the variables of the feature vectors, the suggested distance models in each experiment, the average error rates over each data split, and the overall average error rates.

9.5.1 Experimental results of scenario 1

Let's remember that in scenario 1, the face recognition system considers 6 images for training and 4 images for testing. Within this framework, Figure 43 presents the average weight values (importance of the variables) and the corresponding standard deviations. The standard deviations are depicted as thin lines over the bars portraying the weights. Note that the Fisherfaces space comprehends 14 dimensions since there are 15 classes in the YALE database. Table 17 shows the suggested distance models (values of p) for each experiment under consideration - 10 repetitions over 10 data splits. For the purposes of testing, we considered the averages of the suggested weights, and the values of p that were mostly suggested for each split. Figure 44 depicts the computed average error rates of the 10 splits over 10 repetitions.

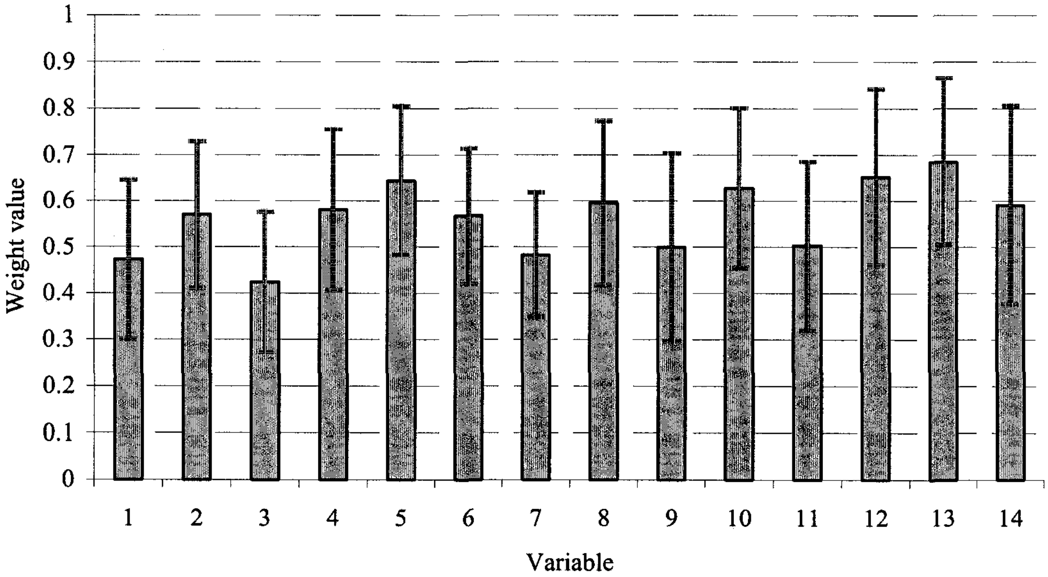


Figure 43 Averages of weights for each variable of Fisher space on YALE (scenario 1)

Table 17 Values of p for the Minkowski distance using YALE (scenario 1)

Repetition	Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Split 8	Split 9	Split 10
1	4	2	8	2	2	6	2	2	10	4
2	6	2	10	2	4	2	2	4	6	2
3	10	6	4	2	2	2	2	4	8	2
4	6	6	6	2	2	8	2	4	4	2
5	8	2	4	2	2	2	2	4	4	2
6	4	8	4	2	2	4	2	2	6	2
7	6	6	8	2	2	2	2	4	6	2
8	8	4	8	2	2	2	2	4	4	4
9	6	8	2	2	2	10	2	2	10	10
10	10	2	10	2	4	6	2	4	10	4

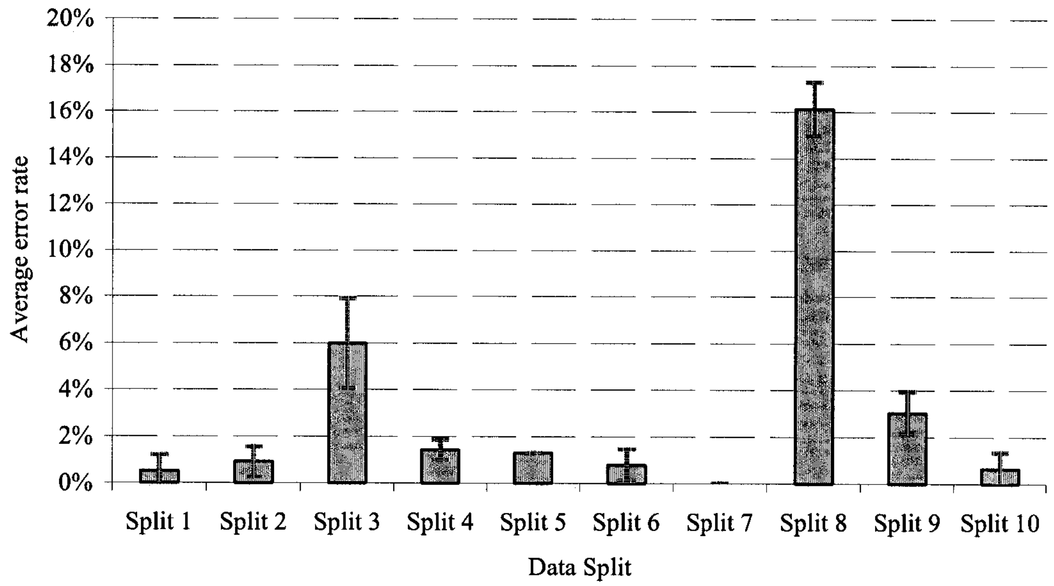


Figure 44 Average error rates over 10 repetitions using YALE (scenario 1)

The overall average error rate and standard deviation are 3.09% ± 0.56%. In other words this is the average of the error rates computed over 100 experiments.

9.5.2 Experimental results of scenario 2

Scenario 2 includes the training, validation, and testing sets. Training and validation sets are used to generate 14 weight values (one for each variable) and 10 parameters p (one for each split). For the purpose of evaluation of the classifiers, we considered the averages of the suggested weights, and the values of p that were mostly suggested for each split. The performance of the face classifiers is evaluated over the test set taking into consideration the suggested weights and parameter p as settings for a weighted Minkowski distance.

Figure 45 presents the averages and standard deviations of the suggested weights over 10 repetitions. Table 18 shows the proposed distance models for each experiment of scenario 2. For comparison purposes, Figure 46 depicts the average error rates and standard deviations computed over the validation sets. It presents the computed results for 10 data splits over 10 repetitions. Lastly Figure 47 illustrates the average error rates and standard deviations computed over the testing sets, the data is organized according to validation split. Please note that in Figure 47 we cannot depict standard deviations since the evaluations were performed only once. This is considering the suggested weights and parameters p. There is no randomness involved in the evaluations over the testing sets.

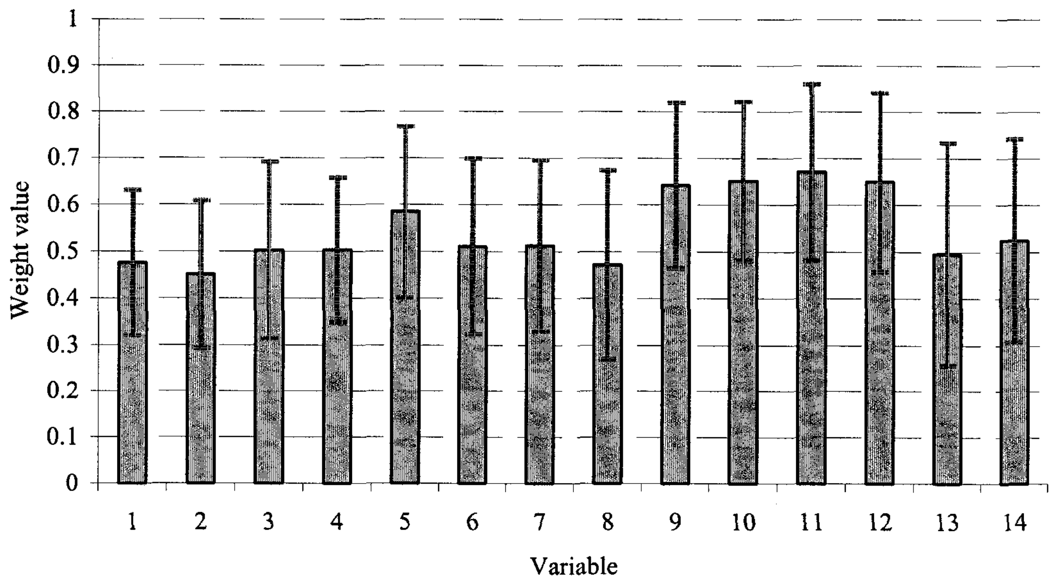


Figure 45 Average weights for each variable of Fisher space on YALE (scenario 2)

Table 18 Values of p for the Minkowski distance using YALE (scenario 2)

Repetition	Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Split 8	Split 9	Split 10
1	2	4	10	2	10	2	2	4	2	4
2	2	10	10	2	2	2	2	2	2	2
3	4	4	10	4	10	2	4	4	2	2
4	2	4	10	2	10	2	4	6	2	2
5	4	4	10	4	10	4	4	4	2	2
6	4	4	10	4	2	4	2	4	2	2
7	2	4	10	8	4	4	4	2	2	2
8	4	4	10	2	10	4	4	4	2	4
9	4	2	10	2	10	2	4	4	2	4
10	4	4	10	4	8	2	2	4	2	2

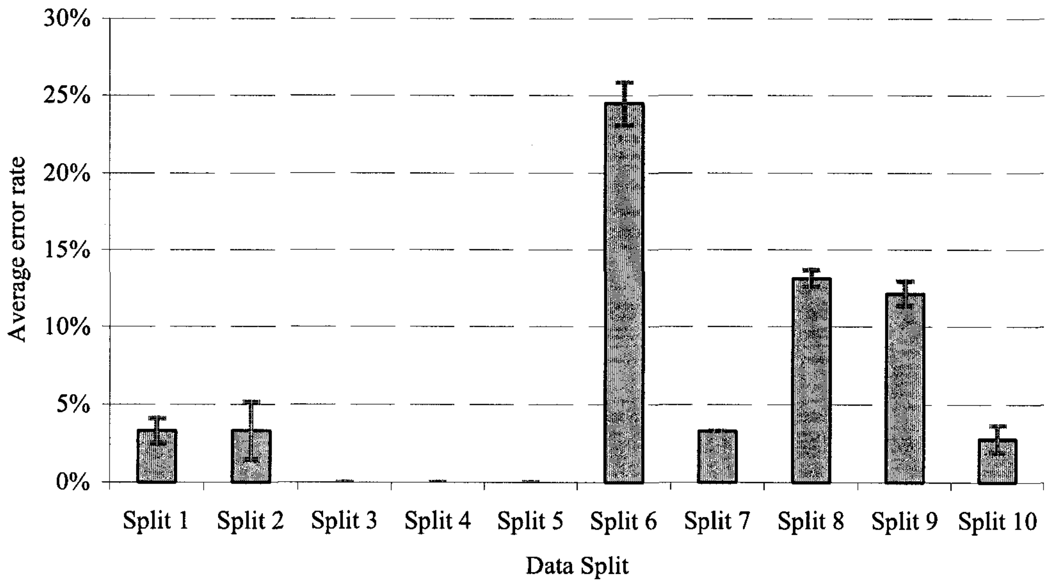


Figure 46 Average error rates over 10 repetitions using YALE (scenario 2 - validation)

The overall average error rate and standard deviation computed over the validation sets are 6.26% ± 0.65%.

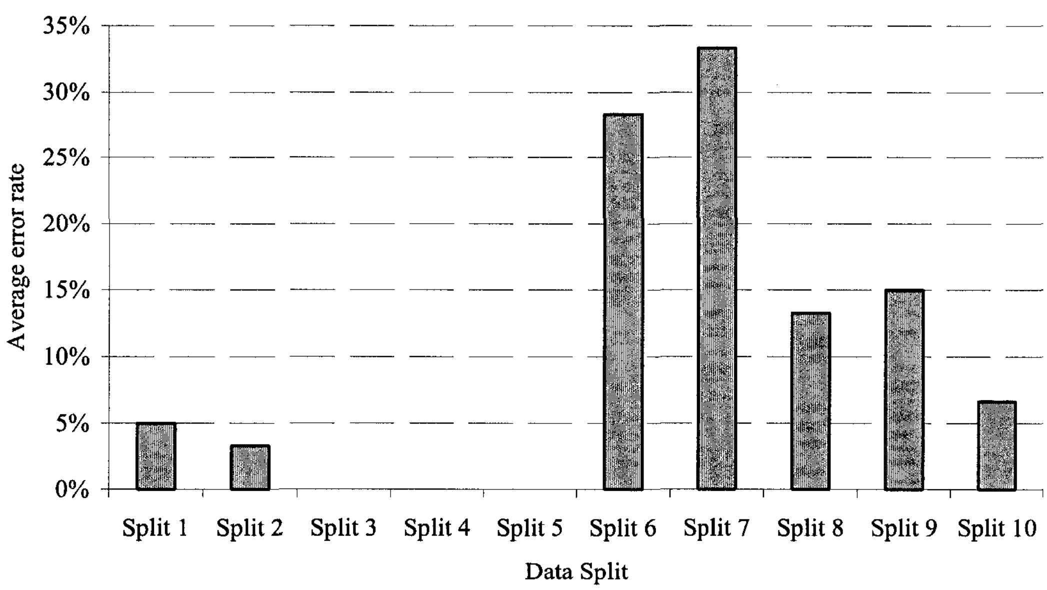


Figure 47 Average error rates over 10 repetitions using YALE (scenario 2 - testing)

The overall average error rate and standard deviation computed over the testing sets are 10.50%, and 11.99% respectively.

Based on the Experimental findings, we can observe an evident improvement of the weighted Minkowski distance over the Fisherfaces classifier, as much as about 8% in scenario 1. In some data splits the error rates even dropped to 0%. In general the GA agrees to a solution for all three repetitions, and a tendency can be observed in the proposed weight values. As for the suggested values of p for the Minkowski distance, the GA proposed Euclidean distance (p=2) as the most suitable parameter most of the times. However, some other values were also suggested.

Some of the differences between the weights values can be attributed to the fact that different values of p were considered in the weighted Minkowski distance, therefore the "importance" of some dimensions may change depending on the values of p (model distance under consideration).

9.6 Conclusion

This chapter has presented an investigation of an evolution-driven approach to face classification. The investigation consisted of an extensive suite of experiments involving Fisherfaces features and GA. The scope of this investigation aims at revealing the importance of the variables forming the feature vectors, and at acquiring suitable distance models for face classification. The importance of the variables is specified by a vector of weights, where each element of the vector is a scalar associated to a particular variable. The distance model is specified by a parameter p of the Minkowski family of distances. Classification is achieved via the nearest neighbor classifier and a weighed Minkowski distance.

There are two scenarios taken into consideration in this study:

- *Scenario 1.* We include 6 and 4 images for training and testing sets correspondingly.
- *Scenario 2.* We include 4, 4, and 3 images for training, validation, and testing sets respectively.

The experimental findings of both scenarios suggest that classification performance can be improved by taking into consideration the importance of the variables involved and a particular distance model. The overall average error rate considering scenario 1 is 3.09. This is the lowest average error rate we have reported for that particular scenario throughout this thesis. As for scenario 2, we report an overall average error rate of 10.50%, which is the lowest we have reported for any single classifier.

The suggested distance models vary according to the data split taken into consideration. In Table 17 and Table 18 we observe that values such as 2 and 4 are generally repeated. However, other distance models are also proposed.

The importance of the variables, as reported by the weight values of Figure 43 and Figure 45, does not show much variability. The values of the weights fall within a narrow region in the middle of the charts. It is interesting then that with such small variations in weights the classification performance improves significantly. The suggested distance model is a factor to consider when suggesting the weights. In other words, different importance may be attributed to each variable depending on the similarity measurement (distance model) under consideration. This explains some of the variations on the suggested weight values.

The reported average error rates in each data split present some significant variations, particularly for splits 6, 7, and 8; see Figure 44, Figure 46, and Figure 47. After careful examination we found out that those splits included more images with variations of lighting in the testing sets.

The computational cost of the presented approach is much higher than that of any other method presented in this thesis. Each experiment of scenario 2 required approximately 25.16 hours for completion, this is considering 1000 generations and a population size of 500 chromosomes. Evidently improvement in classification comes with a price tag.

Chapter 10

CONCLUSIONS

This thesis has presented extensive research in the area of face recognition. Important issues that currently affect face recognition systems have been identified and explained. Aspects such as illumination conditions, facial expressions, environmental disturbances, image quality, image enhancement and transformations, and size of databases have been included. In addition, the small sample size problem – a frequent scenario in face recognition, has been contemplated throughout the investigations.

This research focuses on some well-identified topics of investigation that require scrutiny. The research activities are oriented towards expanding our knowledge in face recognition, and ultimately towards offering design recommendations and guidelines for the improvement of this technology. Various significant research objectives were drawn and fulfilled:

- To assess some meaningful linear and non-linear methods for the construction of feature spaces with respect to classification performance.
- To evaluate the impact of image deterioration on the performance of face classifiers.
- To reveal the effect of image transformations in face classification.
- To investigate the influence of image resolution on the performance of classifiers.
- To inspect methods of aggregation of classifiers in face recognition.
- To explore evolutionary optimization in face recognition towards improving classification.

A number of methods for constructing meaningful feature spaces have been studied and investigated, such as Eigenfaces, Fisherfaces, kernel-PCA, and Isomap. The investigations involve a number of approaches aimed at improving classification rates. They include a modular approach for face recognition, aggregation of classifiers, and evolution-inspired algorithms.

With the outlined objectives in mind, this research has enumerated and carried out various lines of research. Particularly this thesis has delivered:

- *An extensive investigation on the performance of two common face classifiers in the presence of deteriorated visual information.* Two major face classifiers were taken into consideration, namely Eigenfaces and Fisherfaces. The investigations revealed and quantified the relationship that exists between the performance of the classifiers with respect to anticipated types and levels of noise. The contributions derived from this line of research are crucial for the design of adequate systems considering particular environments. Three distortion models are included: Gaussian, salt and pepper, and blurring. We offered guidelines and recommendations for system design in terms of suitable distance models, classifier performance, and levels of noise. We found that training the systems with corrupted and uncorrupted images lead to improved classification. The improvement is attributed to two factors: 1.- there is a better representation of each individual in the feature space since there are more images in the training sets. And 2.- the training environment already contains distorted images which are similar to those in the testing sets. In this investigation we were able to report a low error rate of 1.5%, which to our knowledge is the lowest ever reported using the FERET database.

- *An investigation on a modular approach to face recognition.* The modular approach aims at overcoming the negative effects of illumination and facial expressions. The modular character of the approach allows a given method to focus in localized areas of the faces. In this investigation we consider Eigenfaces as a vehicle that leads to representative features. The behavior of the modular approach is compared to that of Eigenfaces. The investigation contemplates a various image partitions in order to reveal the behavior of the modular approach. Important comments relevant for system design are offered, particularly from the point of view of computational costs. The findings of this investigation suggest that the modular approach indeed overcomes problems of illumination and gesture. We reported an improvement of about 12% over Eigenfaces. As a general trend, the upper regions of the faces portray important discriminatory information compared to the bottom sections. It is possible that mouth gestures introduce high variances even for images of the same individual, hence making that area less relevant for classification. On the other hand, the modular approach comes with higher computational costs.
- *A comprehensive study on the effect of image resolution and image transformations in face recognition.* The study contemplates various methods for constructing feature spaces capable of sound dimensionality reduction, such as Eigenfaces, Fisherfaces, Isomap, and kernel-PCA. A number of suitable parameters are explored in each method. In addition, some meaningful image transformations were included in the investigations, namely histogram equalization and edge detection. The findings portray important practical implications to systems' design. This study revealed the tradeoff that exists between classifier performance vis-à-vis image resolutions and image transformations. The fact that this study takes into consideration several methods for dimensionality reduction, image transformations, and image resolutions makes this contribution unique. With respect to image resolution, we offer evidence suggesting that high-quality images do not necessarily lead to better recognition. In fact, a relatively small number of pixels are required to achieve acceptable performance. We also show that contrast enhancement and edge detection help improve accuracy. The improvements range from about 6% to 18% depending on the database, image resolution, image transformation, and feature space.
- *An assessment of aggregation of classifiers with respect image transformations.* The feature spaces under consideration were constructed by means of Eigenfaces, Fisherfaces, kernel-PCA, and Isomap. Various parameters were explored in each method. The aggregation architecture took into consideration features constructed from contrast enhancement, edge detection, and images without manipulation. The aggregation methods include majority voting and Bayesian product rule. The performance of single classifiers is compared to that of combined architectures. Our previous research on the impact of image resolution helped us select and adequate image resolution without sacrificing accuracy. We support our findings with the statistical t-test. Useful findings and design recommendations are offered. In this

investigation we achieved better performance by combining features via the Bayesian product rule. In general, the lowest error rates were delivered by kernel-PCA with polynomial kernel of second order, and by Fisher-Isomap. We were able to report an average error rate of about 4% with images of YALE, and about 7% with images of FERET. Overall, we have confirmed the hypothesis that contrast enhancement and edge detection provide additional discriminatory information useful for classification.

- *An exploration on the importance of the variables forming the feature vectors in a given space.* This research is based on the hypothesis that the variables forming the feature vectors portray particular discriminatory information useful for classification. The variables are commonly selected by ranking them according to their particular variances. However, the variances may not reflect their true "importance" when it comes to distinguishing individuals. We presented a method that reveals the importance of each variable from the classification point of view. In addition, the presented method suggests a suitable similarity measure for differentiating individuals in a given feature space. The similarity models are derived from the Minkowski family of distances. This approach relays on Evolutionary optimization by Genetic Algorithms to accomplish the task. Experimental evidence supports the usefulness of this approach. The outcomes of this research are appealing. We were able to report an average error rate of 3.09% for the best case, which is the lowest we have achieved in our investigations without image distortions. In some cases this method achieved 0% error rate in some data splits. In general, the GA is able to suggest similar distance models and "importance" for each variable after many repetitions. Despite the encouraging results, the method portrays major drawbacks in terms of computational costs and susceptibility to dimensionality of feature space. In terms of computational costs, this method is much more expensive than any other reported in this thesis, in some cases the training phase lasted for about 25 hours. With respect to size of feature space, if the feature space is too broad then this method fails to improve classification. However, it does not perform worst than nearest neighbor classifier using the same feature space.

This thesis has proposed and explored various approaches to face recognition. In each case the advantages and limitations of the designs have been highlighted. The success of a face recognition system depends greatly on the environmental conditions in which it is to operate. The contributions of this thesis include useful design recommendations and guidelines for a number of realistic scenarios. The research findings are sustained on an extensive suite of carefully designed experiments, on some useful methods for constructing feature spaces, and on meaningful face databases regarded as benchmarks.

For further work we suggest two main streams of research:

- *Exploring other algorithms for classification.* It would be important to explore the performance of the proposed architectures and feature spaces considering other algorithms of classification. So far we have investigated on the performance of face recognition taking into consideration the nearest neighbor classification rule, nonetheless other methods may decrees error rates. May we suggest non-linear classifiers such as Neural Networks (NN), or fuzzy classifiers such as Fuzzy C-Means (FCM), and perhaps even k-nearest neighbor with $K>1$. The data we have presented in this thesis may serve as baseline for future work. An assessment of performance considering other classifiers may deliver important findings for face recognition.

- *Exploring on other methods for constructing feature spaces.* So far we have investigated extensively on Eigenfaces, Fisherfaces, Isomap, and kernel-PCA. It would be illustrative and useful to investigate on other alternatives for representing images. Based on our findings, we would recommend non-linear approaches, perhaps Local Linear Embedding (LLE) [85]or Curvilinear Distance Analysis (CDA) [54], to mention just a few.

Our experimentation provided encouraging results regarding the modular approach for face recognition. Further investigation on this approach may deliver important design recommendations. As starting point, we would suggest considering Fisherfaces, Isomap, and kernel-PCA as suitable feature spaces for the modular representation of the images. Surely we would also encourage involving other feature spaces in this investigation. In our investigations of the modular approach we divided the images into equal-sized sub-images. However, other sizes may be investigated.

In terms of image transformations, we have delivered interesting results taking into account contrast enhancement and edge detection. Some variations of these transformations may provide more distinctive information for classification, for instance edge detection with thresholding. Binary edge images may depict more robust distinctive information when dealing with lighting conditions.

Our research showed the benefits and drawbacks of Genetic Algorithms in face recognition. The most predominant negative issue is the time required for training, perhaps some variations in the presented method could deliver a faster architecture. We would suggest doing training in a series of successive steps. This is training with a small number of individuals at first, and once GA has proposed some weights we could include more images for training. The searching of the new weights would take less time since some approximations would have already been made during the previous training. The intention would be to reduce the overall training time. One possible complication is to fall into local minima. However, increasing the probability of mutation may reduce this effect. Other improvements can be made regarding software implementation. Our software was developed to run on single processors, it would be very beneficial to develop multi-thread applications for research on evolution-inspired methods. Computing time may be reduced up to 100 times by using the machines we have access to at the Intelligent System Laboratory (ISL).

This thesis reports on the impact of image deterioration on face classifiers, namely Eigenfaces and Fisherfaces. We included three distortion models, such as Gaussian, blurring, and salt and pepper noise. Our investigations can be extended to other classifiers, such as Isomap and kernel-PCA. At this point it may also be beneficial to include combined types of noise on the images. It would also be appealing to investigate on the behavior of the classifiers including images with different types and levels of distortions in the training sets. The expected effect would be to have a much better representation of each individual in the feature space. This would alleviate the effects of the small sample size problem. The second intrinsic consequence would be a more robust system given that the classifiers would train with corrupted images as well.

Bibliography

[1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 721-732, July 1997.

[2] M. J. Aitkenhead, and A. J. S. McDonald, "A neural network face recognition system", *Engineering Applications of Artificial Intelligence*", Vol. 16, Issue 3, pp. 167-176, April 2003.

[3] S. Arca, P. Campadelli, and R. Lanzarotti, "A face recognition system based on automatically determined facial fiducial points", *Pattern Recognition*, Vol. 39, Issue 3, pp. 432–443, 2006.

[4] S. Arca, P. Campadelli, R. Lanzarotti, and G. A. Lipori, "A face recognition system dealing with expression variant faces", *Proc. of the 18th IEEE Intl. Conf. on Pattern Recognition*, Vol. 2, pp. 1242-1245, 2006.

[5] M. J. Aitkenhead, and A. J. S. McDonald, "A neural network face recognition system", *Engineering Applications of Artificial Intelligence*", Vol. 16, Issue 3, pp. 167-176, April 2003.

[6] M. A. Anjum, M. Y. Javed, "Face recognition vs. image resolution", *Trans. of the First Intl. Conf. on Information and Communication Technologies*, pp. 109-112, 2005.

[7] T. Bäck, and H.-P. Schwefel, "Evolutionary computation: and overview", *Proc. of IEEE Intl. Conf. on Evolutionary Computation*, pp. 20-29, May 1996.

[8] W. A. Barrett, "A survey of face recognition algorithms and testing results", *Conf. Records of the Thirty-First Asilomar Conference on Signals, Systems, and Computers*, Vol. 1, pp. 301-305, November 1997.

[9] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis", *IEEE Trans. on Neural Networks*, Vol. 13, No. 6, pp. 1450-1464, November 2002.

[10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, July 1997.

[11] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.

[12] R. M. Bolle, et al., *Guide to Biometrics*, Springer-Verlag New York, 2004.

[13] S. Cagnoni, A. Poggi, and G. L. Porcari, "A modified modular eigenspace approach to face recognition", *Proc. of the IEEE Intl. Conf. on Image Analysis and Processing*, pp. 490-495, 1999.

[14] W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain", *IEEE Trans. on Systems, Man and Cybernetics, Part B (Cybernetics)*, Vol. 36, No. 2, pp. 458-466, 2006.

[15] J. Choi, S. Lee, C. Lee, and J. Yi, "A real-time face recognition system using multiple mean faces and dual mode Fisherfaces" *Proc. of the IEEE Intl. Symposium on Industrial Electronics*, Vol. 3, pp. 1686 -1689, June 2001.

[16] K. Cios, W. Pedrycz, and R. Swiniarski, *Data mining methods for knowledge discovery*, Kluwer Academic Publishers, Second Printing, 2000.

[17] D. Coltuc, P. Bolon, and J-M. Chassery, "Exact histogram specification", *IEEE Trans. on image processing*, Vol. 15, Issue 5, pp. 1143-1152, 2006.

[18] T. M. Cover, and P. E. Hart, "Nearest neighbor pattern classification", *IEEE Trans. on Information Theory*, Vol. IT-13, pp. 21-27, January 1967.

[19] S. Cruz-Llanas, J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, "A comparative evaluation of global representation-based schemes for face verification", *Proc. of the Intl. Conf. on Image Processing*, Vol. 3, pp. III-905 – III-908, September 2003.

[20] E. R. Dougherty, and C. R. Giardina, *Matrix structured image processing*, Prentice Hall, Englewood Cliffs, NJ, 1987.

[21] R. O. Duda, P. E. Hart, *Pattern classification and scene analysis*, John Wiley & Sons, New York, 1973.

[22] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics*, Vol. 3, pp. 32-57, 1973.

[23] M. P. Ekstrom, *Digital image processing techniques*, Academic Press, 1984.

[24] M. J. Er, S. Wu, J. Lu, and H. Lye Toh, "Face recognition with radial basis function (RBF) neural networks", *IEEE Trans. on Neural Networks*, Vol. 13, No. 3, pp. 697-710, May 2002.

[25] R. A. Fisher, "The use of multiple measurements in taxonomic problems" *Annals of Eugenics,* Vol. 7, pp. 179-188, 1936.

[26] B.J. Frey, A. Colmenarez, and T.S. Huang, "Mixtures of local linear subspaces for face recognition", *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 32-37, June 1998.

[27] Y. Gao, M. K. H. Leung, "Face recognition using line edge map", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, pp. 764-779, 2002.

[28] R. Gottumukkal, "An improved face recognition technique based on modular PCA approach", *Pattern Recognition Letters*, Vol. 25, No. 4, pp. 429-436, March 2004.

[29] B. Guo, K.-M. Lam, K.-H. Lin, and W.-C, Siu, "Human face recognition based on spatially weighted hausdorff distance", *Pattern Recognition Letters*, Vol. 24, Issue 1-3, pp. 499-507, January 2003.

[30] S. Haider, A. Abbas, A. K. Zaidi, "A multi-technique approach for user identification through keystroke dynamics", *IEEE Intl. Conf. on Systems, Man, and Cybernetics*, Vol. 2, pp. 1336-1341, October 2000.

[31] Z. Hong, "Algebraic feature extraction of image for recognition" *Pattern Recognition*, Volume 24, Issue 3, pp. 211-219, 1991.

[32] W. Hsu and M. F. Tenorio, "Software engineering effort models using neural networks", *IEEE Intl. Conf. on Neural Networks*, pp. 1190-1195, 1991.

[33] Y. Ivanov, B. Heisele, and T. Serre, "Using component features for face recognition", *Proc. of the 6th IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 421-426, 2004.

[34] A. K. Jain, "Biometric recognition: how do I know who you are?", *Proc. of the 12th IEEE Conf. on Signal Processing and Communications Applications*, pp. 3-5, April 2004.

[35] A. K. Jain, A. Ross, "Learning user-specific parameters in a multibiometric system", *Proc. of the Intl. Conf. on Image Processing*, Vol. 1, pp. I-57 – I-60, September 2002.

[36] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, No. 1, pp. 4-20, January 2004.

[37] A. K. Jain, S. Pankanti, "A touch of money [biometric authentication systems]", *IEEE Spectrum*, Vol. 43, Issue 7, pp. 22-27, July 2006.

[38] G. Jarillo Alvarado, W. Pedrycz, M. Reformat, K.-C. Kwak, "Eigenfaces versus weighted modular PCA for human face recognition", *Proc. of the Symposium on Human-Centric Computing: HC205*, pp. 19-25, February 2005.

[39] G. Jarillo Alvarado, W. Pedrycz, M. Reformat, K.-C. Kwak, "Deterioration of visual information in face classification using Eigenfaces and Fisherfaces", *Machine Vision and Applications*, Vol. 17, No. 1, pp. 68-82, March 2006.

[40] T. Kam Ho, J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 66-75, January 1994.

[41] T. Kanade, *Picture processing system by computer complex and recognition of human faces*, Dept. of Information Science, Kyoto University, November 1973.

[42] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the Sobel operator", *IEEE journal of solid-state* circuits, Vol. 23, Issue 2, pp. 358-367, 1998.

[43] M. A. U. Khan, M. T. Ibrahim, M. K. Khan, and M. A. Khan, "Cross correlation measure for decision fusion among multiple face classifiers", *Proc. of the IEEE Symposium on Emerging Technologies*, pp. 126-131, 2005.

[44] T. Khoshgoftaar, *et al.*, "Using neural networks to predict software faults during testing", *IEEE Trans. on Reliability*, Vol. 45, No. 3, pp. 456-462, September 1996.

[45] H.-C. Kim, D. Kim, S. Y. Bang, S.-Y. Lee, "Face recognition using the second-order mixture-of-Eigenfaces method", *Pattern Recognition*, Vol. 37, Issue 2, pp. 339-349, February 2004.

[46] M. Kirby and L. Sirovich, "Application of karhunen-love procedure for characterization of human faces", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, Issue 1, pp. 103-108, January 1990.

[47] J. Kittler, "Combining classifiers: a theoretical framework", *Pattern Analysis and Applications*, Vol. 1, No. 1, pp. 18-27, March 1998.

[48] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp. 226-239, March 1998.

[49] A. Z. Kouzani, F. He, and K. Sammut, "Multiresolution eigenface-components", *Proc. of the IEEE Region 10 Annual Conf. on Speech and Image Technologies for Computing and Telecommunications*, Vol. 1, pp. 353-356, 1997.

[50] E. P. Kukula, S. J. Elliott, "Securing a restricted site-biometric authentication at entry point", *Proc. IEEE 37th Annual Intl. Carnahan Conf. on Security Technology*, pp. 435-438, October 2003.

[51] R. Kumar, S. Rai, J. L. Trahan, "Neural-network techniques for software-quality evaluation", *IEEE Proc. of the Annual Reliability and Maintainability Symposium*, pp. 155-161, January 1998.

[52] L. I. Kuncheva, "Fuzzy versus non-fuzzy in combining classifiers designed by boosting", *IEEE Trans. on Fuzzy Systems*, Vol. 11, No. 6, pp. 729-741, 2003.

[53] K. N. Leach, "A survey paper on independent component analysis", *Proc. of the Thirty-Fourth Southeastern Symposium on System Theory*, pp. 239-242, March 2002.

[54] J. A. Lee, A. Lendasse, and N. Donckers, and M. Verleysen, "A robust nonlinear projection method", *Proc. of the 8th European Symposium on Artificial Neural Networks*, pp. 13-20, April 2000.

[55] J. A. Lee, A. Lendasse, and M. Verleysen, "Curvilinear distance analysis versus Isomap", *Proc. of the 10th European Symposium on Artificial Neural Networks*, pp. 185-192, April 2002.

[56] J. A. Lee, A. Lendasse, and M. Verleysen, "Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis", *Neurocomputing*, Vol. 57, pp. 49-76, 2004.

[57] R. Liao, and S. Z. Li, "Face recognition based on multiple facial features", *Proc. of the Fourth IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 239-244, March 2000.

[58] K.-H. Lin, K.-M. Lam, and W.C. Siu, "Spatially eigen-weighted hausdorff distance for human face recognition" *Pattern Recognition*, Vol. 36, Issue 8, pp. 1827-1834, August 2003.

[59] C. Liu; H. Wechsler, "Evolution of optimal projection axes (OPA) for face recognition", *Proc. of the Third IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp 282-287, April 1998.

[60] C. Liu and H. Wechsler, "A unified Bayesian framework for face recognition", *Proc. of the IEEE Intl. Conf. on Image Processing*, Vol. 1, pp. 151-155, October 1998.

[61] C. Liu and H. Wechsler, "Enhanced fisher linear discriminant model for face recognition" *IEEE Trans. on Image Processing*, Vol. 11, No. 4, pp. 467-476, April 2002.

[62] C. Liu and H. Wechsler, "Independent component analysis of gabor features for face recognition", *IEEE Trans. on Neural Networks*, Vol. 14, No. 4, pp. 919-928, July 2003.

[63] N. Liu, H. Wang, "Feature extraction with genetic algorithms based nonlinear principal component analysis for face recognition", *Proc. of the 18th IEEE Intl. Conf. on Pattern Recognition*, Vol. 3, pp. 461-464, August 2006.

[64] C. G. Looney, *Pattern recognition using neural networks: theory and algorithms for engineers and scientists*, Oxford University Press, New York, 1997.

[65] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms", *IEEE Trans. on Neural Networks*, Vol. 14, No. 1, pp. 117-126, January 2003.

[66] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms", *IEEE Trans. on Neural Networks*, Vol. 14, No. 1, pp. 195-200, January 2003.

[67] D. Maio, *et al.*, "FVC2000: fingerprint verification competition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 3, pp. 402-412, March 2002.

[68] C. Mair, *et al*, "An investigation of machine learning based prediction systems", *The journal of systems and software*, Vol. 53, pp. 23-29, 2000.

[69] P. McGuire and G. M. T. D'Eleuterio, "Eigenpaxels and a neural network approach to image classification", *IEEE Trans. on Neural Networks*, Vol. 12, No. 3, pp. 625-635, May 2001.

[70] B.A. McLindin, M.A. Butavicius, and K. Meaney, "Gallery image effects on facial recognition systems", *4th Conf. on Video/Image Processing and Multimedia Communications*, Vol. 2, pp. 455-460, 2003.

[71] P. Melin, C. Gonzalez, and D. Bravo, "Modular neural networks with fuzzy Sugeno integral for pattern recognition", *Annual Meeting of the North American Fuzzy Information Processing Society*, pp. 602-606, June 2005.

[72] D. C. Montgomery, and G. C Runger, *Applied Statistics and Probability for Engineers*, John Wiley & Sons, New York, 1999.

[73] X. Mu, P. Watta, and M.H. Hassoun, "Combining local similarity measures: summing, voting, and weighted voting", *IEEE Intl. Conf. on Systems, Man and Cybernetics*, Vol. 3, pp. 2661-2666, 2005.

[74] M. Y. Nam, B. Bayarsaikhan, and P. K Rhee, "Face Recognition Using Correlation Between Illuminant Context", *Lecture Notes in Computer Science*, Vol. 4029 pp. 833-840, 2006.

[75] W. Pedrycz, "Algorithms of fuzzy clustering with partial supervision", *Pattern Recognition Letters*, Vol. 3, Issue 1, pp. 13-20, January 1985.

[76] H. Peng, G. Rong, Z. Bian, "Automated face recognition using adaptive subspace method", *IEEE Intl. Conf. on Systems, Man, and Cybernetics*, Vol. 1, pp. 88-92, October 1996.

[77] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition", *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[78] A. Pentland and T. Choudhury, "Face recognition for smart environments", *IEEE Computer Journal*, Vol. 33, Issue 2, pp. 50-55, February 2000.

[79] V. Perlibakas, "Distance measure for PCA-based face recognition", *Pattern Recognition Letters*, Vol. 25, Issue 6, pp. 711-724, April 2004.

[80] T. V. Pham, M. Worring, A. Smeulders, "Face detection by aggregated Bayesian network classifiers", *Pattern Recognition Letters*, Vol. 23, Issue 4, pp. 451-561, 2002.

[81] P. J. Phillips, *et al.*, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing J.*, Vol. 16, No. 5, pp 295-306, 1998.

[82] D. Reisfeld and Y. Yeshurun, "Robust detection of facial features by generalized symmetry", *Proc. of the 11th Intl. Conf. on Pattern Recognition*, Vol. 1, pp. 117-120, September 1992.

[83] A. Reza and H. Yan, "Human Face Image Recognition: An Evidence Aggregation Approach", *Computer Vision and Image Understanding*, Vol. 17, No. 2, pp. 213-230, August 1998.

[84] B.D. Ripley, *Pattern recognition and neural networks*, Cambridge University Press, New York, 1996.

[85] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science*, Vol. 290, No. 5500, pp. 2323-2326, December 2000.

[86] F.S. Samaria, and A.C. Harter, "Parameterisation of a stochastic model for human face identification", *Proc. of the Second IEEE Workshop on Applications of Computer Vision*, pp. 138-142, December 1994.

[87] S. Sanderson and J. H. Erbetta, "Authentication for secure environments based on iris scanning technology", *IEE Colloquium on Visual Biometrics*, pp. 8/1-8/7, March 2000.

[88] C. Sanderson, K. K. Paliwal, "Noise compensation in a milti-modal verification system", *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 157-160, May 2001.

[89] C. Sanderson, K. K. Paliwal, "Noise resistant audio-visual verification via structural constraints", *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. V-716-V-719, April 2003.

[90] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlineal component analysis as a kernel eigenvalue problem", *Neural Computation*, Vol. 10, pp. 1299-1319, 1998.

[91] Y. Shu, T. Tan, and Y. Wang, "Biometric personal identification based on handwriting", *Proc. of the 15th Intl. Conf. on Pattern Recognition*, Vol. 2, pp. 797-800, September 2000.

[92] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces", *J. Opt. Soc. Amer., A*, Vol. 4, Issue 3, pp. 519-524, 1987.

[93] I. Stainvas and N. Intrator, "Blurred face recognition via a hybrid neural architecture", *Proc. of the 15th IEEE Intl. Conf. on Pattern Recognition*, Vol. 2, pp. 805-808, September 2000.

[94] B. Takács, and H. Wechsler, "Face recognition using binary image metrics", *Proc. of the Third IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 294-299, 1998.

[95] J. B. Tenenbaumm V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, Vol. 290, pp. 2319-2323, December 2000.

[96] Y. Tian, T. Tan, Y. Wang, and Y. Fang, "Do singular values contain adequate information for face recognition?", *Patten Recognition*, Vol. 36, Issue 3, pp. 649-655, March 2003.

[97] M. Tistarelli and E. Grosso, "Active face recognition with a hybrid approach", *Pattern Recognition Letters*, Vol. 18, Issue 9, pp. 933-946, September 1997.

[98] M. Turk and A. Pentland, "Face recognition using Eigenfaces," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[99] G. K. Venayagamoorthy, V. Moonasar, and K. Sandrasegaran, "Voice recognition using neural networks", *Proc. of the 1998 South African Symposium on Communication and Signal Processing*, pp. 29-32, September 1998.

[100] D. Whitley, "An overview of evolutionary algorithms: practical issues and common pitfalls", *Information and Software Technology*", Vol. 43, Issue 14, pp. 817-831, December 2001.

[101] X. Xie, K.-M. Lam, "An efficient illumination compensation scheme for face recognition", *8th Intl. Conf. on Control, Automation, Robotics and Vision*, Vol. 2, pp. 1240-1243, 2004.

[102] L. Xu, A. Krzyżak, and C.Y. Suen, "Methods for combining multiple classifiers and their applications to handwriting recognition", *Trans. on Systems, Man, and Cybernetics*, Vol. 22, No. 3, pp. 418-435, 1992.

[103] Y. Xu, B. Li, and B. Wang, "Face recognition by fast independent component analysis and genetic algorithm", *Proc. of the Fourth IEEE Intl. Conf. on Computer and Information Technology*, pp. 194-198, September 2004.

[104] Z. Yankun, L. Chongqing, "Face recognition using kernel principal component analysis and genetic algorithms", *Proc. of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 337-343, September 2002.

[105] A. Yilmaz, M. Gökmen, "Eigenhill vs. eigenedge", *Proc. of the 15th Intl. Conf. on Pattern Recognition*, Vol. 2, pp. 827-830, 2000.

[106] H. Yu and J. Yang, "A direct lda algorithm for high-dimensional data with application to face recognition", *Pattern Recognition*, Vol. 34, pp. 2067-2070, 2001.

[107] B. L. Zhang, H. Zhang, and S. S. Ge, " Face recognition by applying wavelet subband representation and kernel associative memory", *IEEE Trans. on Neural Networks*, Vol. 15, No. 1, pp. 166-177, January 2004.

[108] W. Zhang, *et al.*, "Information fusion in face identification", *Proc. of the 17th Intl. Conf. on Pattern Recognition*, Vol. 3, pp. 950-953, 2004.

[109] W.-S. Zheng, J.-H. Lai, and P. C. Yuen, "GA-fisher: a new LDA-based face recognition algorithm with selection of principal components", *IEEE Trans. on Systems, Man, and Cybernetics – Part B*, Vol. 35, No. 5, pp. 1065-1078, October 2005.

[110] W. Zhao, R. Chellappa, "Robust image based face recognition", *Proc. of the IEEE Intl. Conf. on Image Processing*, pp. 41-44, Vol. 1, September 2000.

[111] W. Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant analysis of principal components for face recognition", *Proc. of the Third IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 336-341, April 1998.

[112] Federal Trade Commission (FTC), "Identity Theft Survey Report", Prepared by Synovate (Aegis Group plc), 2003.

[113] http://www.cognitec-systems.de/index.html (January 2007)

[114] http://www.cybula.com/ (January 2007)

[115] http://www.dreammirh.com/english/main.html (January 2007)

[116] http://www.identix.com/company/comp_overview.html (January 2007)

[117] http://www.neurotechnologija.com/index.html (January 2007)

[118] http://www.omniperception.com/ (January 2007)

[119] http://www.biometrics.org/html/introduction.html (January 2007)

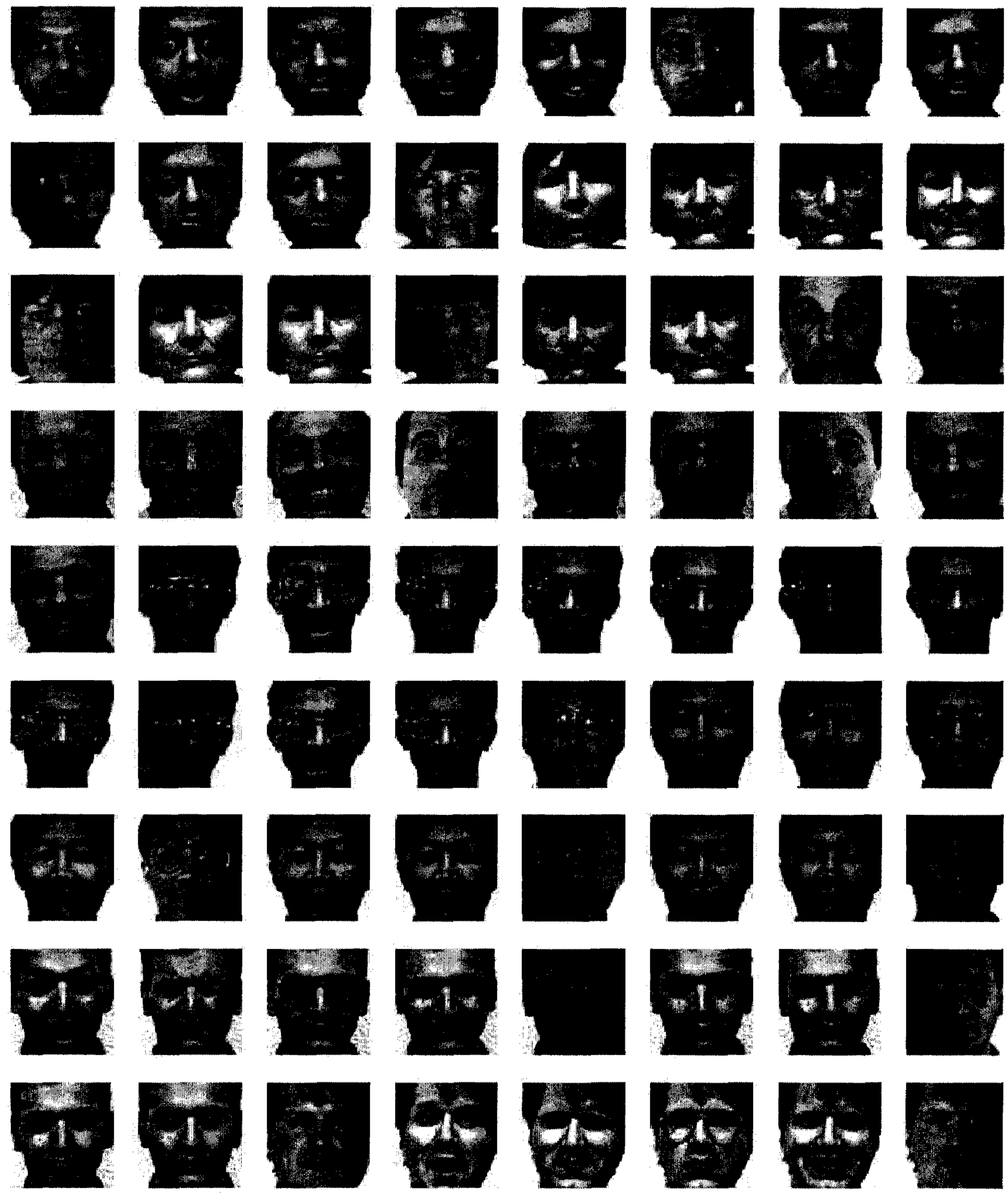
[120] http://cvc.yale.edu/projects/yalefaces/yalefaces.html (January 2007)

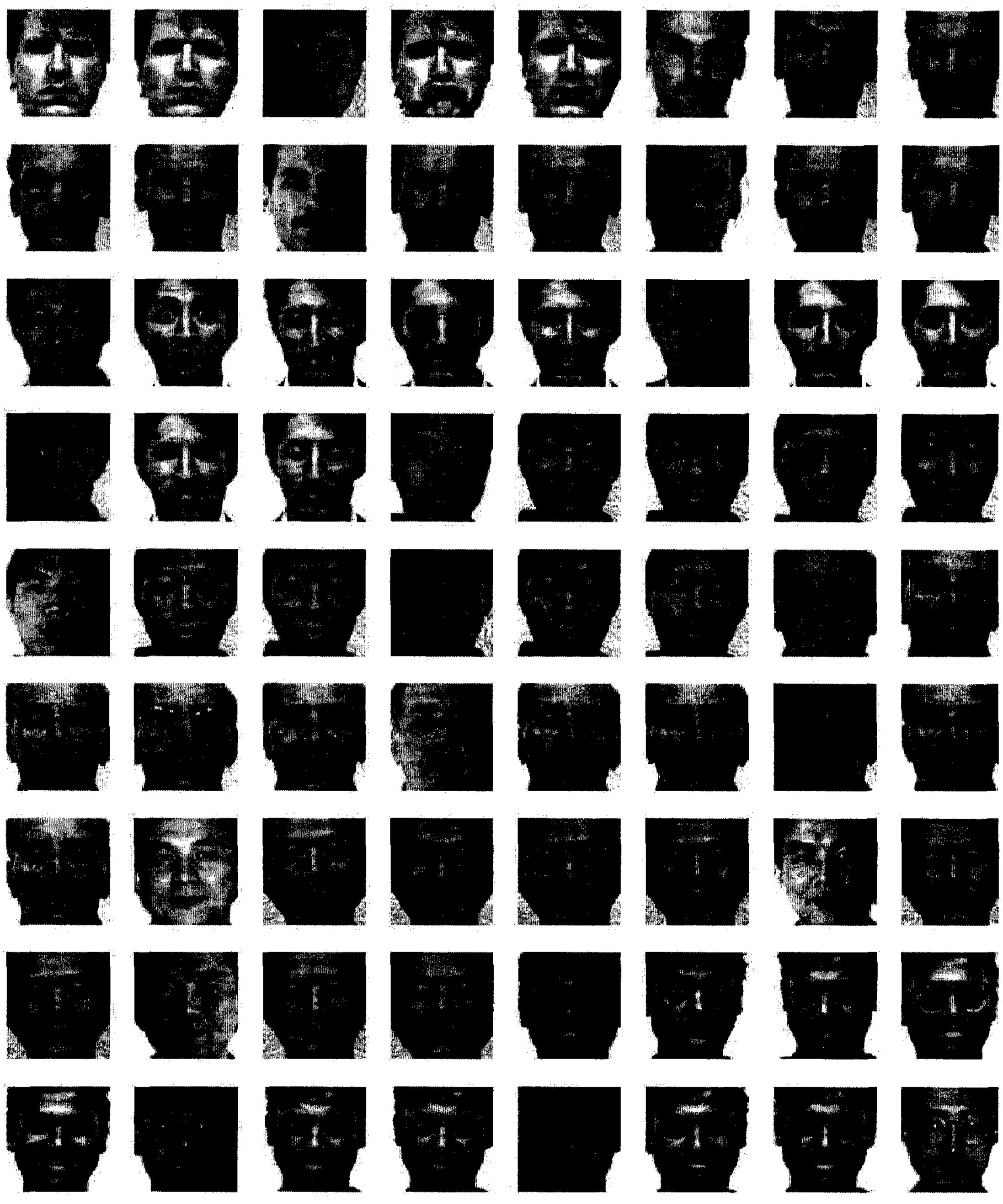
[121] http://www.ibia.org/biometrics/ (January 2007)

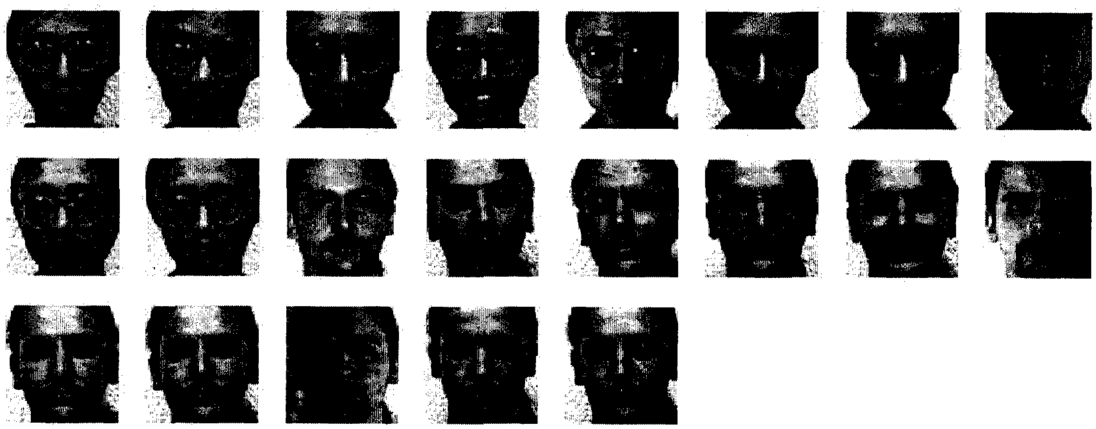
[122] http://www.econ.iastate.edu/tesfatsi/holland.GAIntro.htm (January 2007)

IMAGES OF THE FERET AND YALE DATABASES

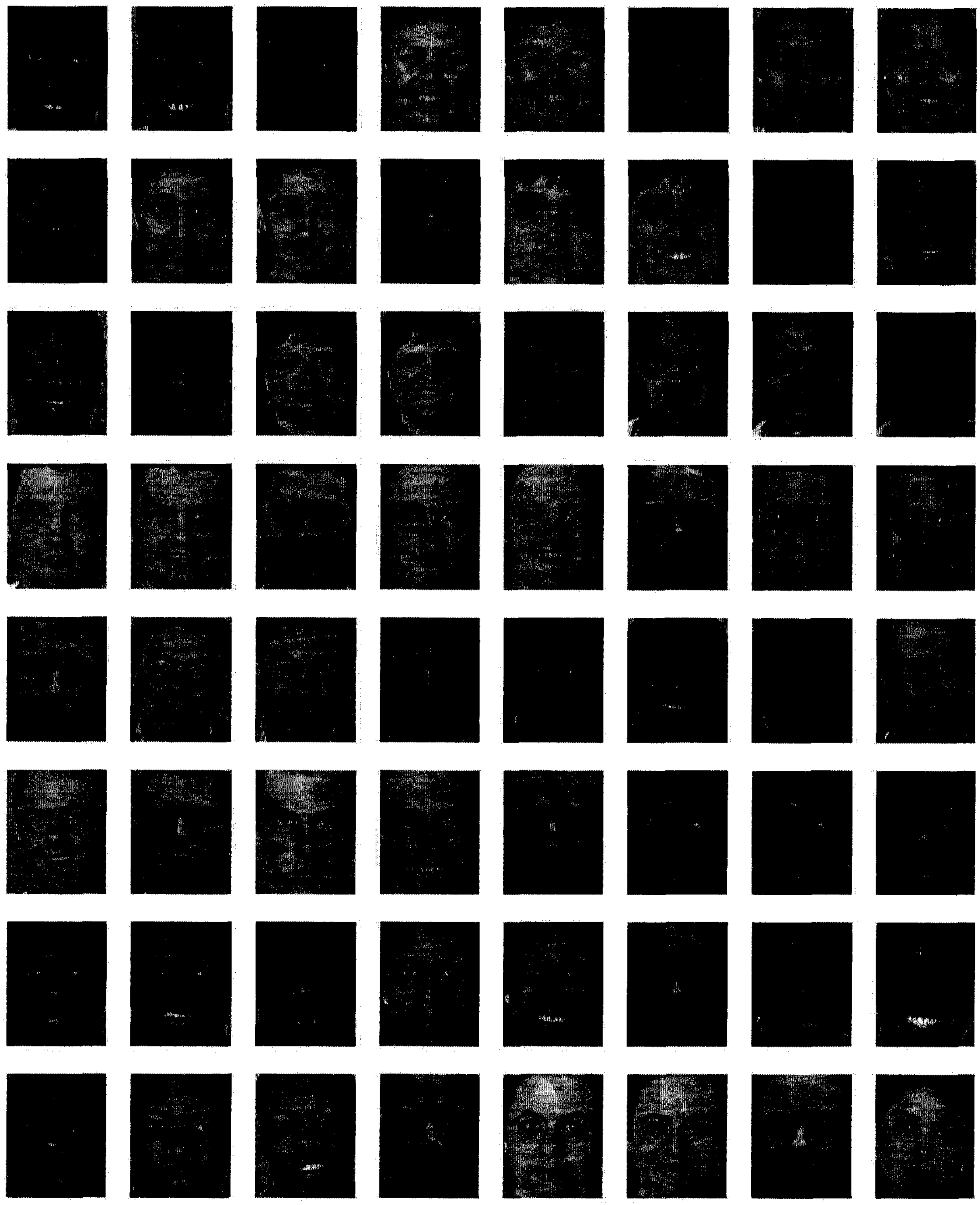
This appendix shows the image databases used in our investigations. The following photos depict the images of the YALE database.

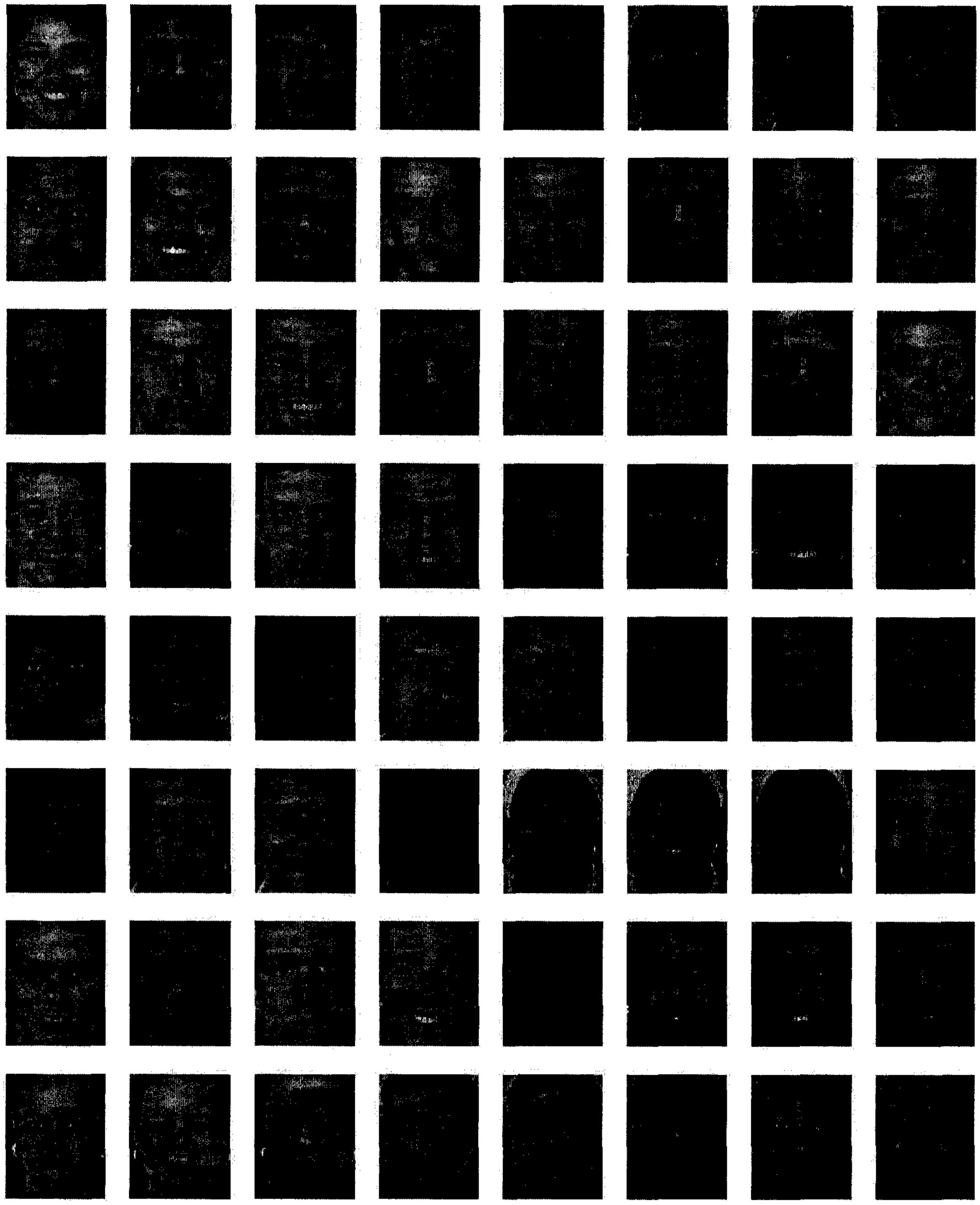


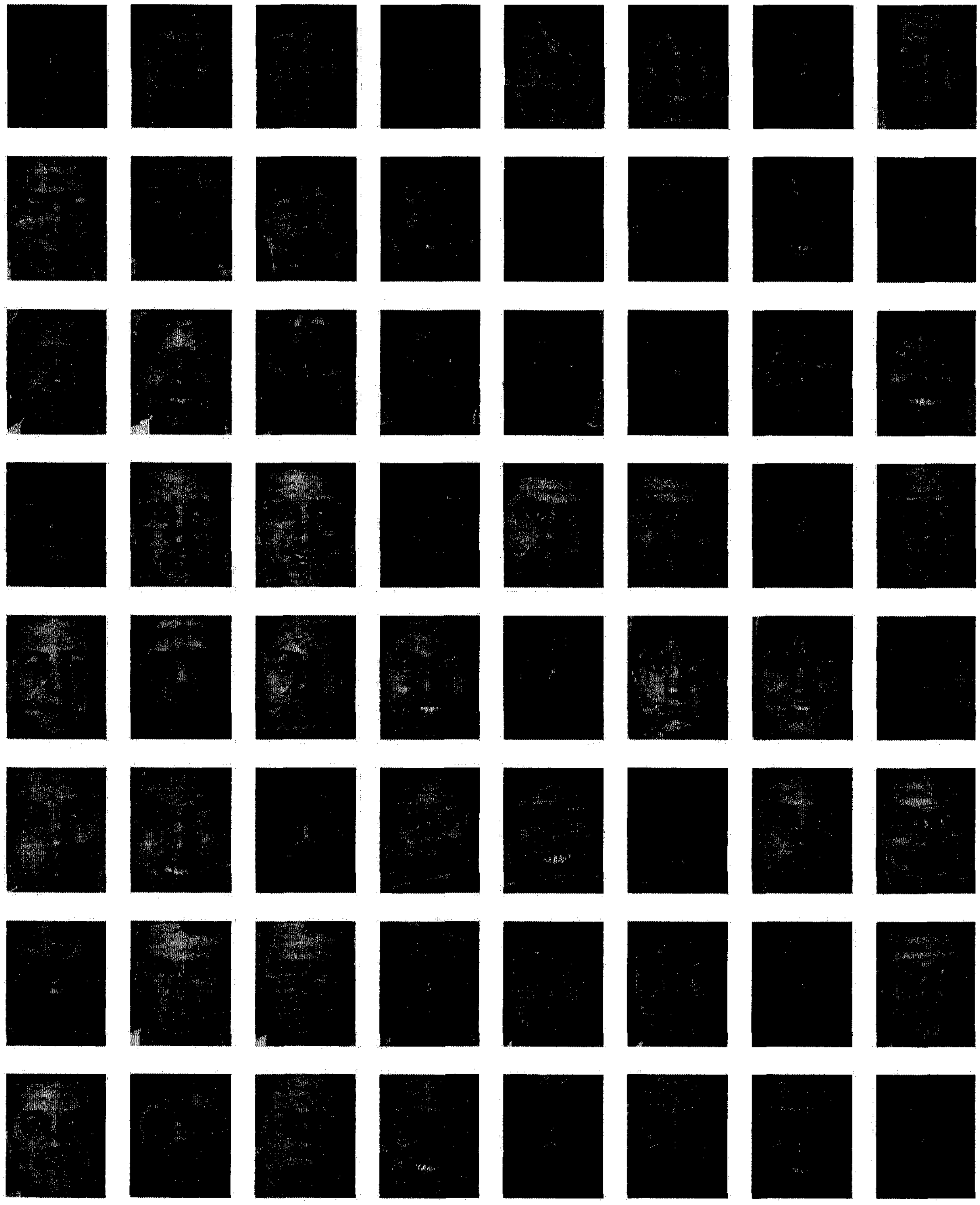


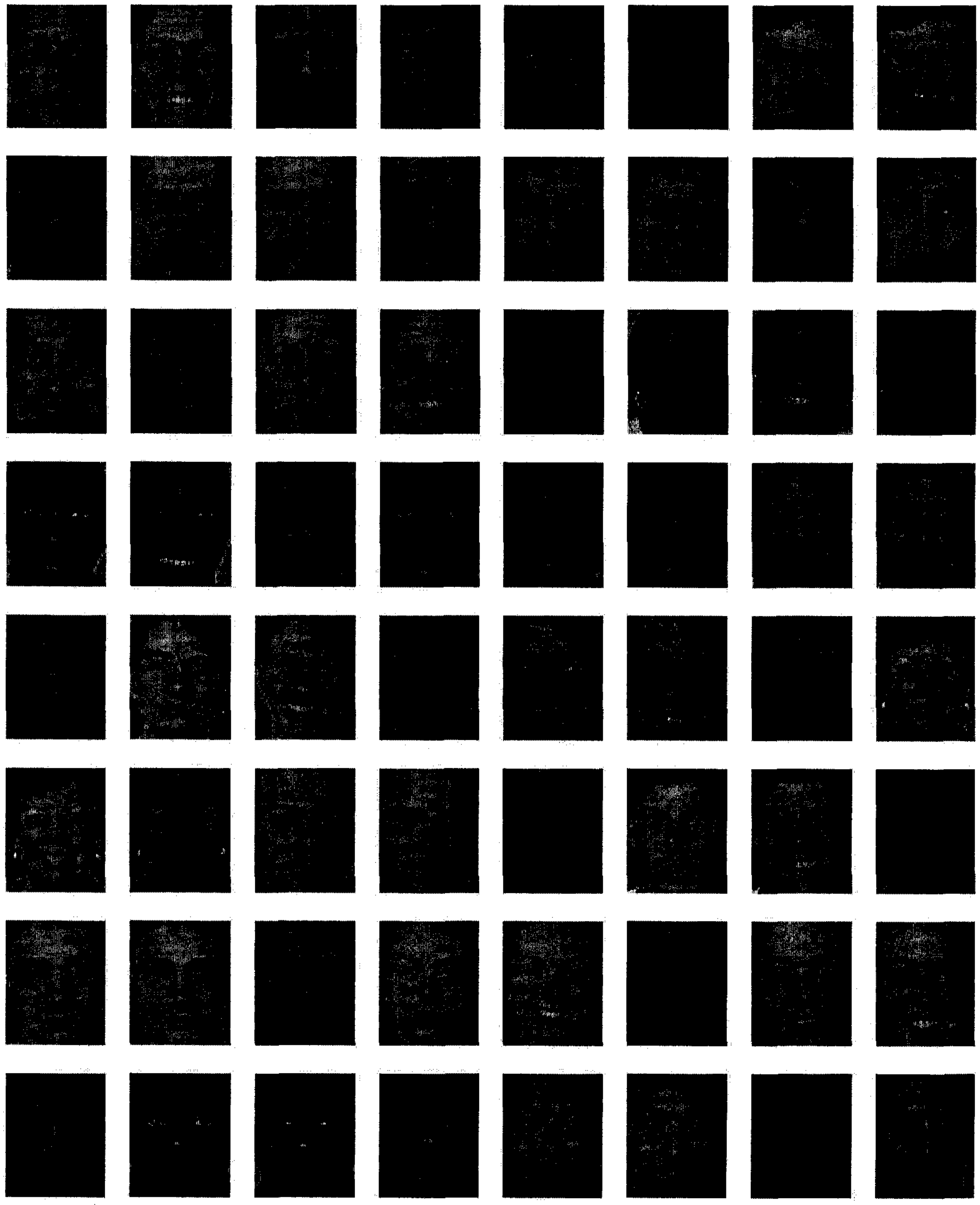


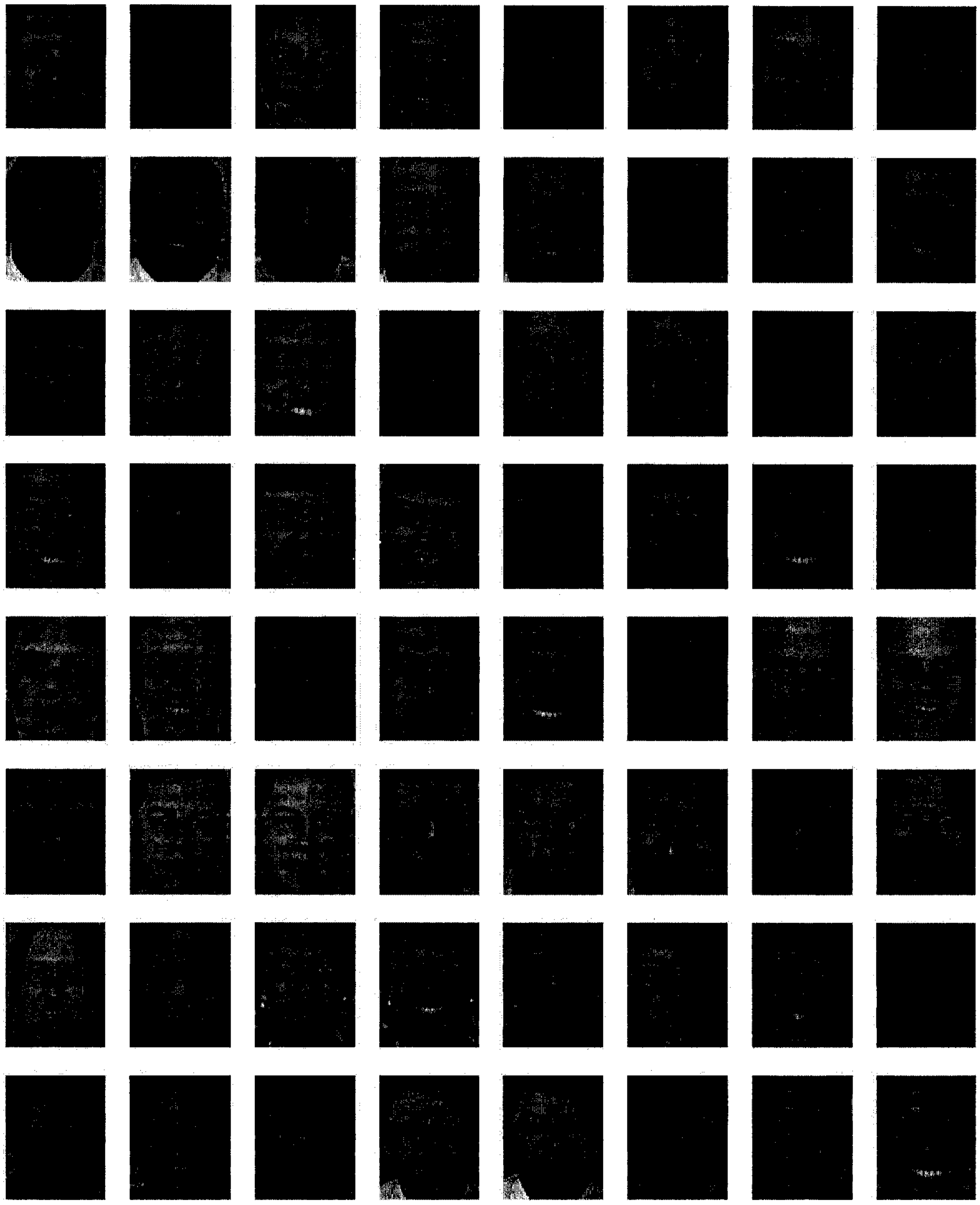
The following photos depict the images of the FERET database.

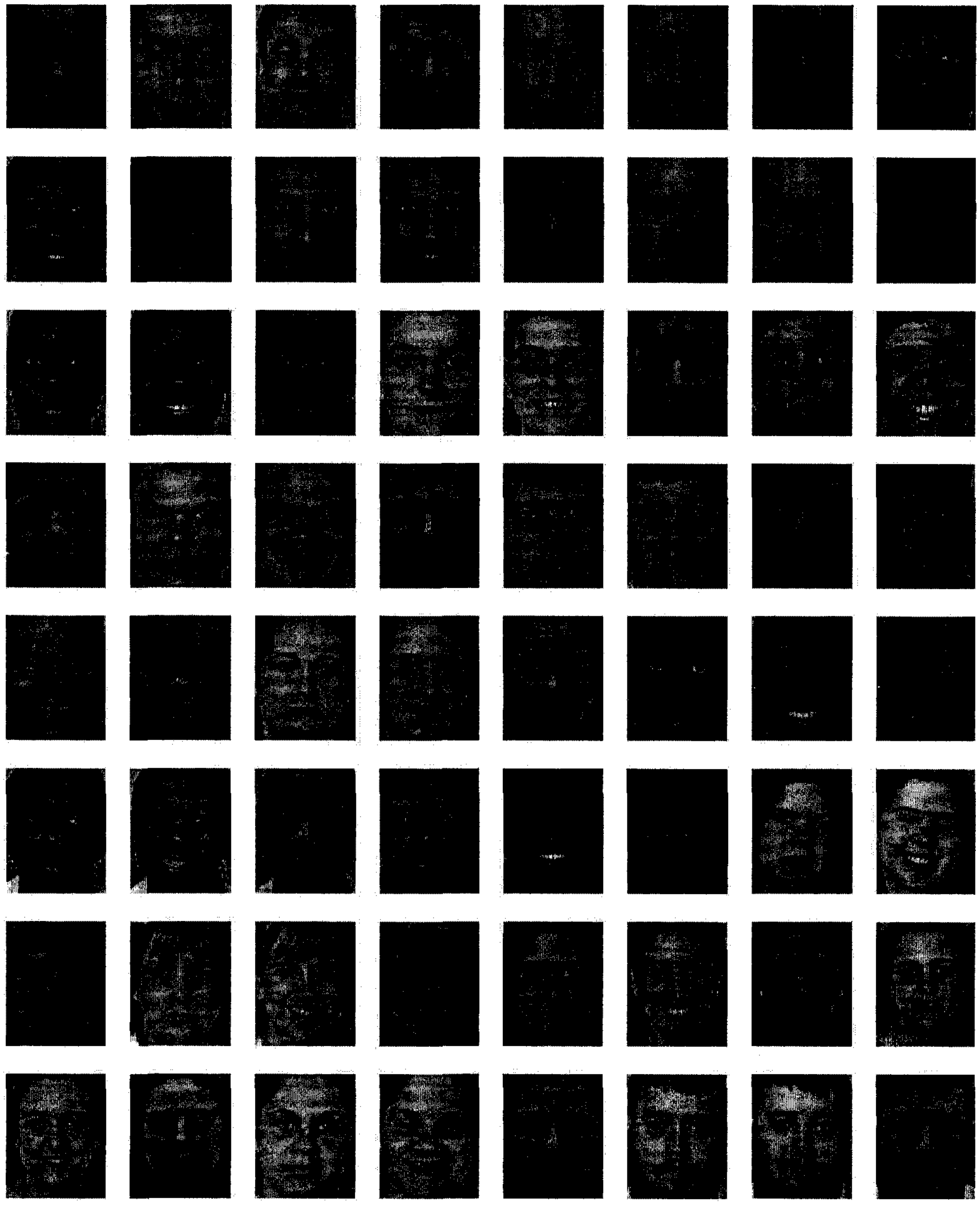


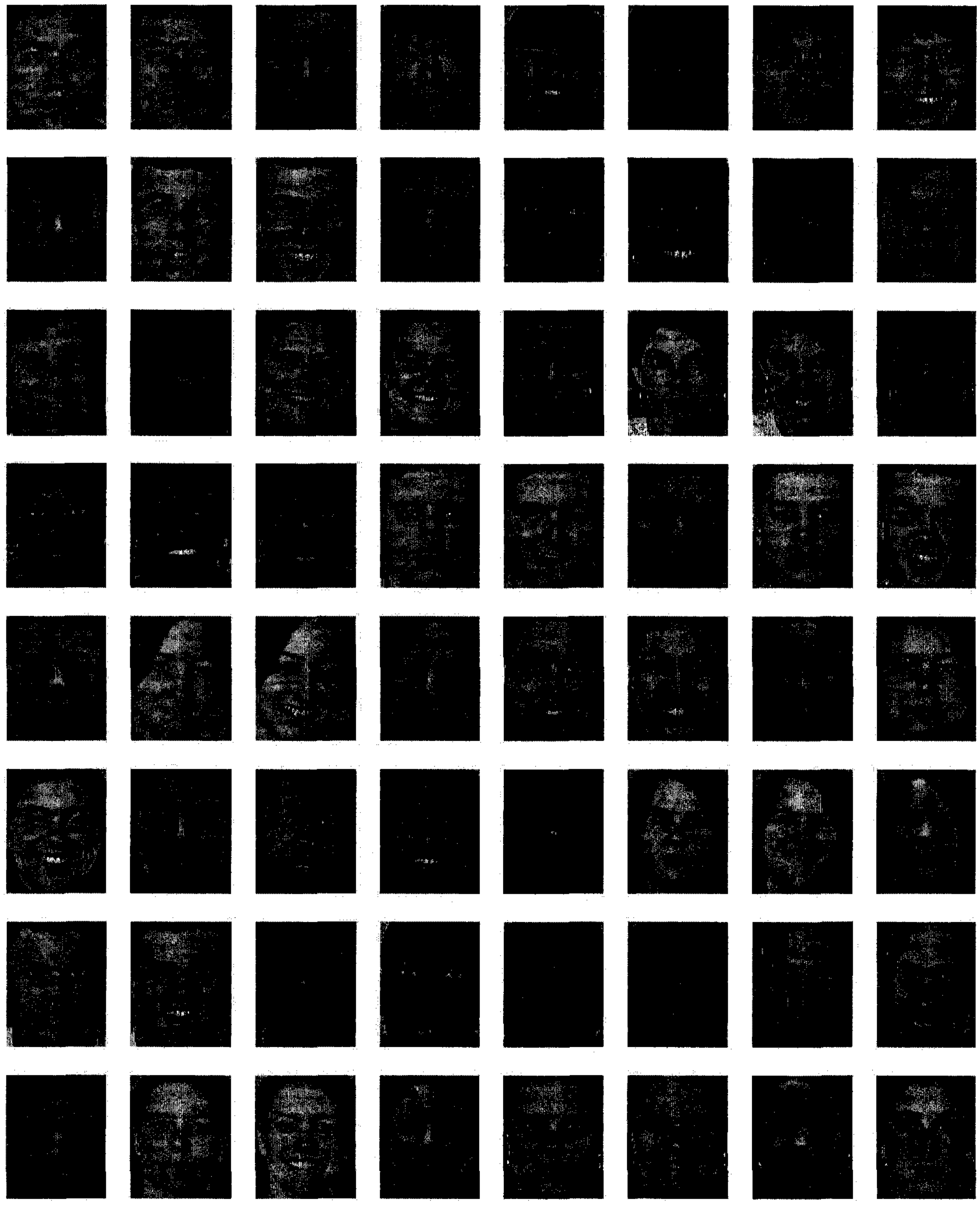


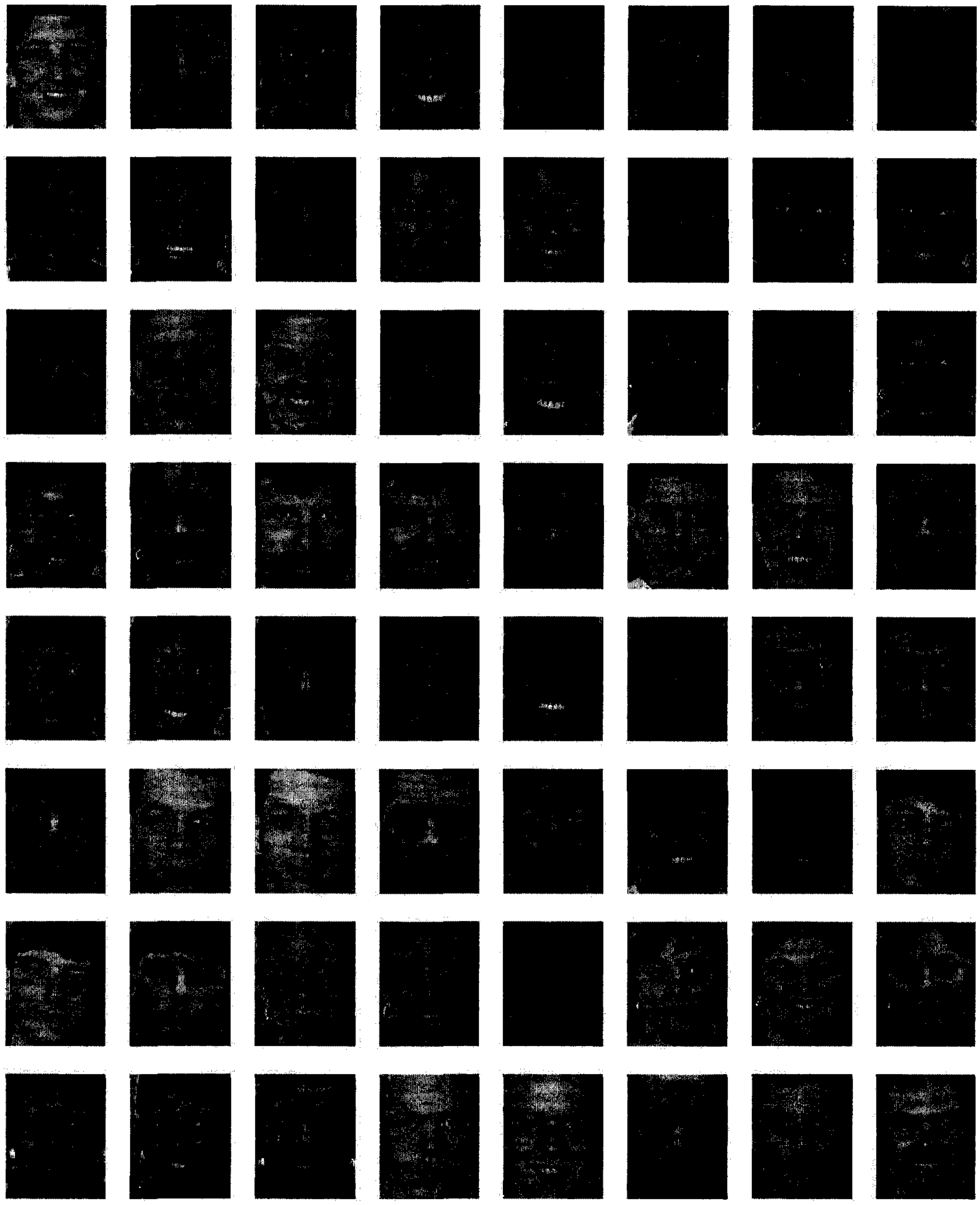


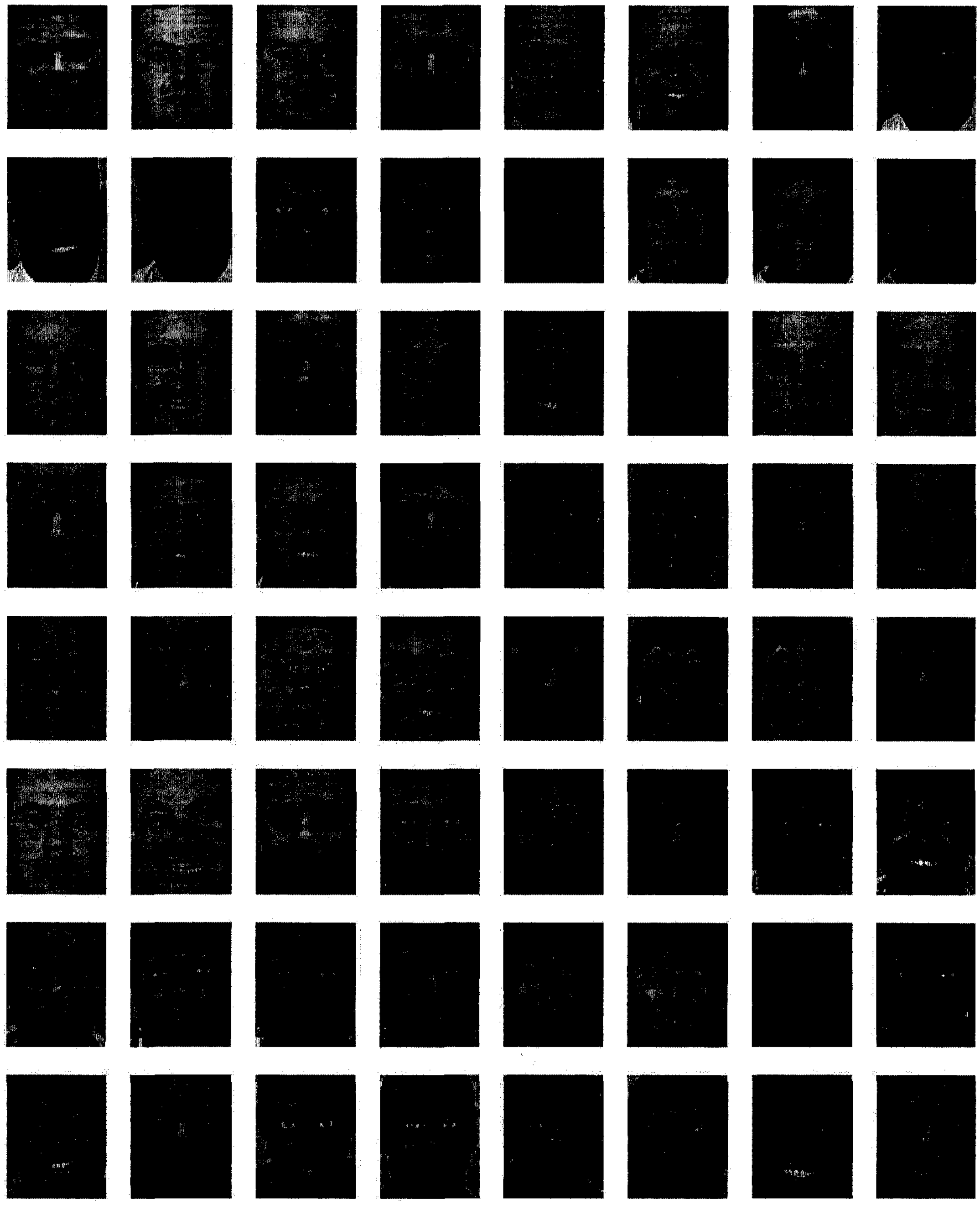


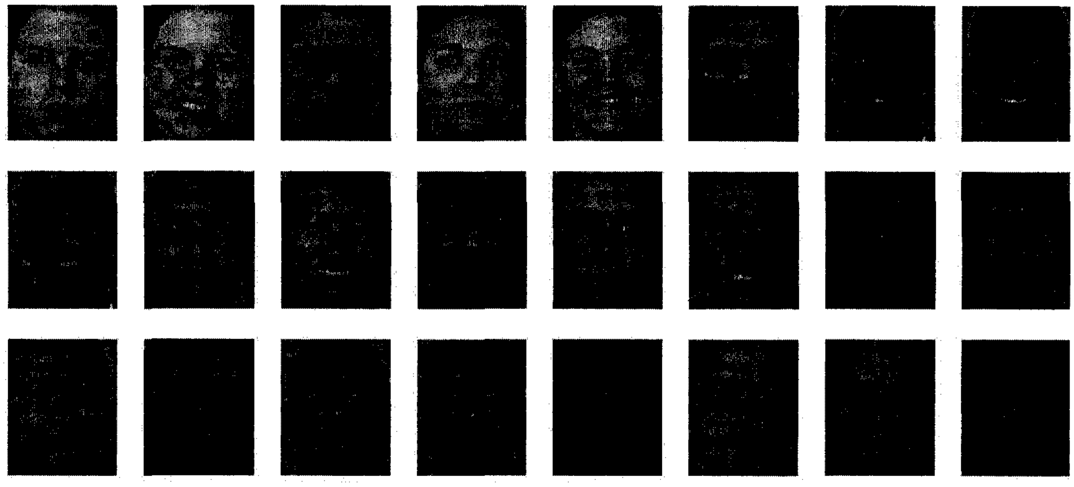












Appendix B

PERFORMANCE OF CLASSIFIERS UNDER VARIOUS IMAGE RESOLUTIONS

This appendix presents a summary of the performance of various classifiers with respect to image resolution. The image distributions for the training, validation, and testing sets are presented in Table 1 located in section 3.1. The experimental results depicted in this appendix were computed taking into consideration the training and testing sets only, the validation sets were not required for this investigation.

The performance of the classifiers is portrayed in a variety of charts. They depict the performance of the classifiers taking into account images with no transformation, contrast-enhanced images, and edge images. The data is organized according to feature space and image quality (horizontal and vertical pixels). The error rates are depicted by bars in each instance. The standard deviations are portrayed as thin lines overlapped with the error rates. As for the experimentation with YALE, Figure 48 depicts the average error rates and standard deviations of the Eigenfaces and Fisherfaces classifiers. Figure 49 shows the performance of the Isomap approach, it includes Eigen-Isomap and Fisher-Isomap. Figure 50 portrays the performance of kernel-PCA with three kernel models, namely polynomial, Gaussian, and sigmoid.

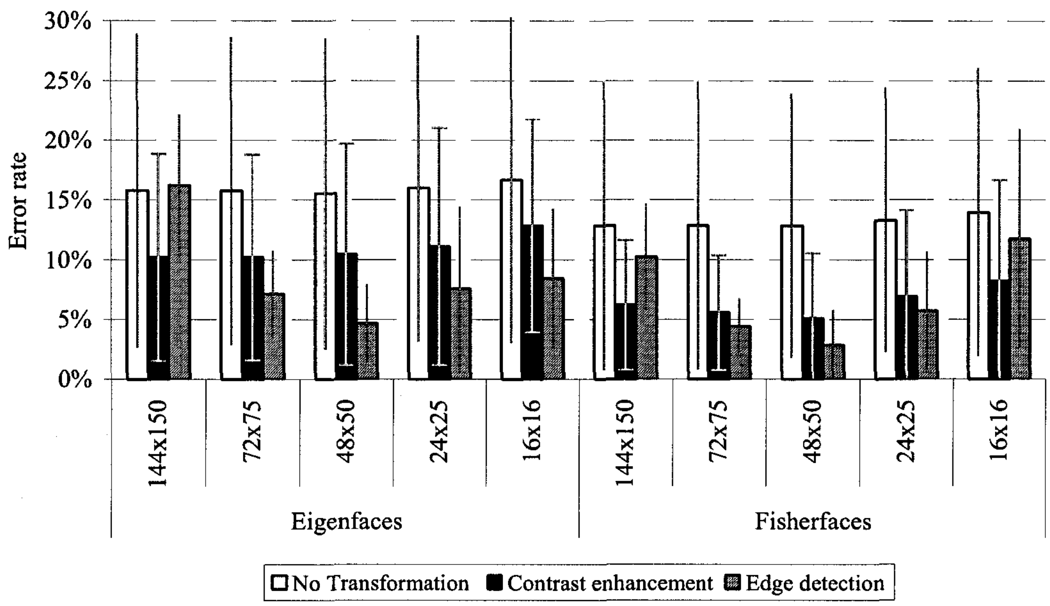


Figure 48 Performance of Eigenfaces and Fisherfaces using images YALE

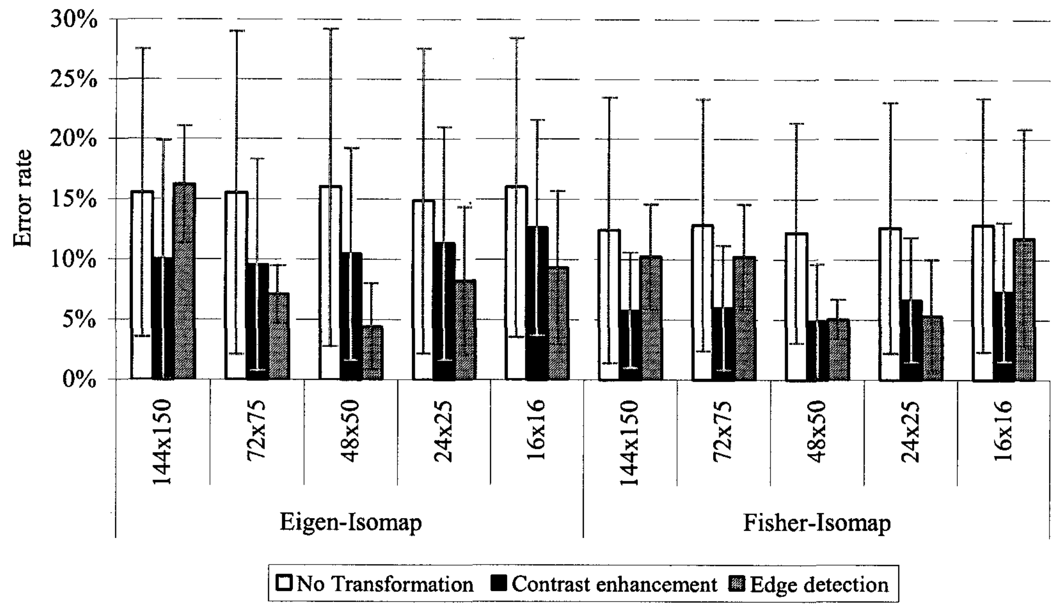


Figure 49 Performance of Eigen-Isomap and Fisher-Isomap using YALE

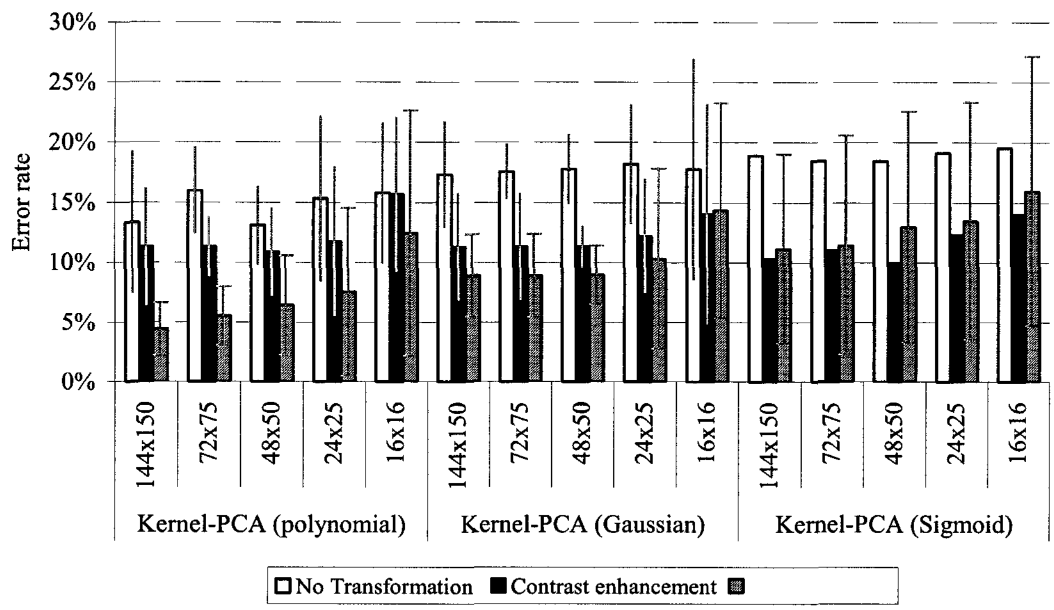


Figure 50 Performance of kernel-PCA using YALE

In regard to the experimentation with FERET, Figure 51 presents the performance of Eigenfaces and Fisherfaces classifiers. Figure 52 depicts the average error rates of Eigen-Isomap and Fisher-Isomap. Lastly Figure 53 portrays the behavior of kernel-PCA with three kernel models. The plots include the performance of the classifiers taking into consideration images with no transformation, contrast-enhanced images, and edge images.

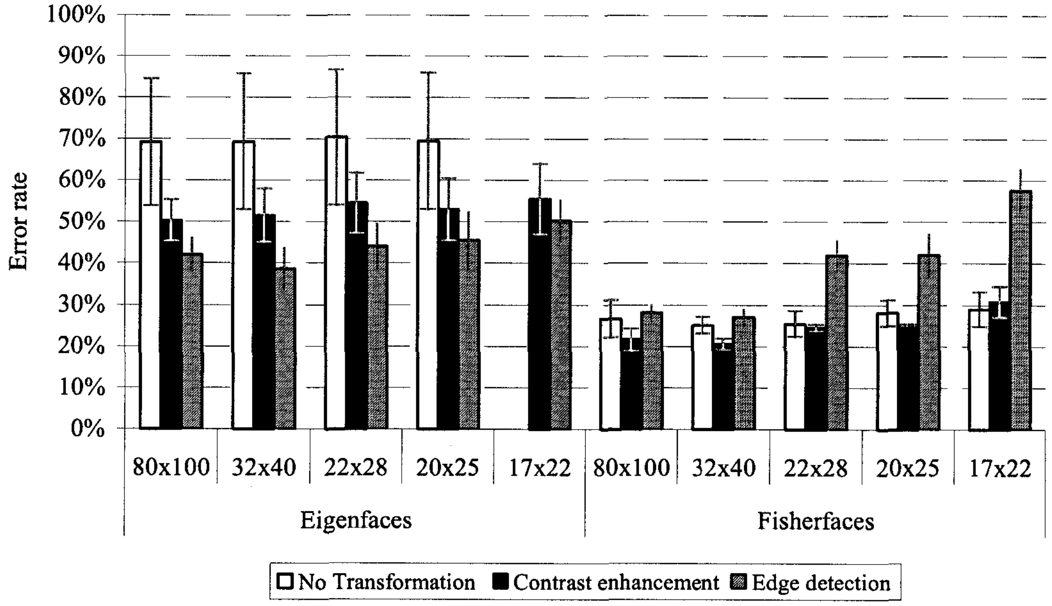


Figure 51 Performance of Eigenfaces and Fisherfaces using FERET

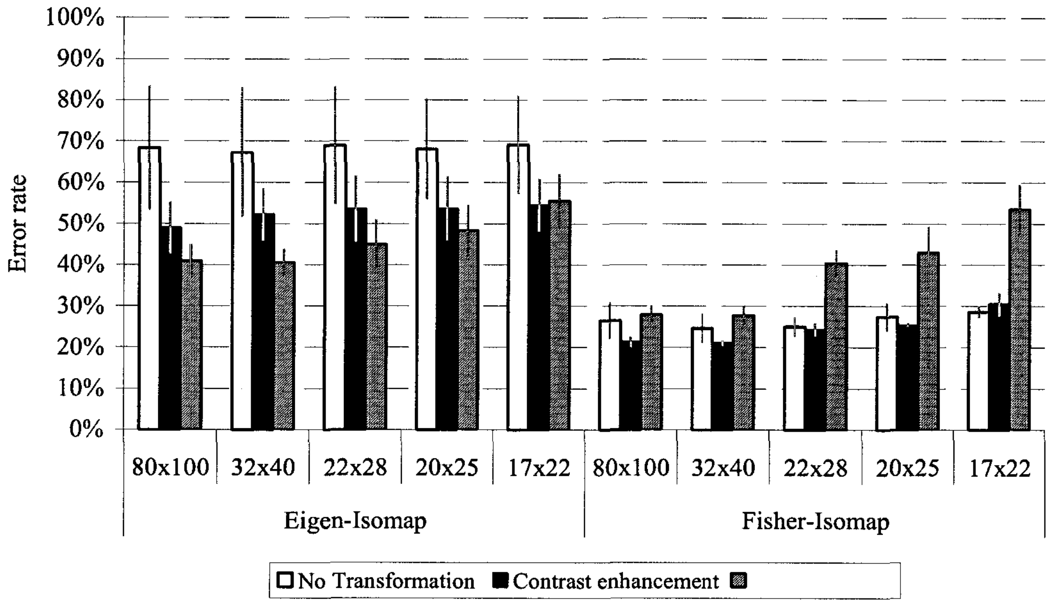


Figure 52 Performance of Eigen-Isomap and Fisher-Isomap FERET

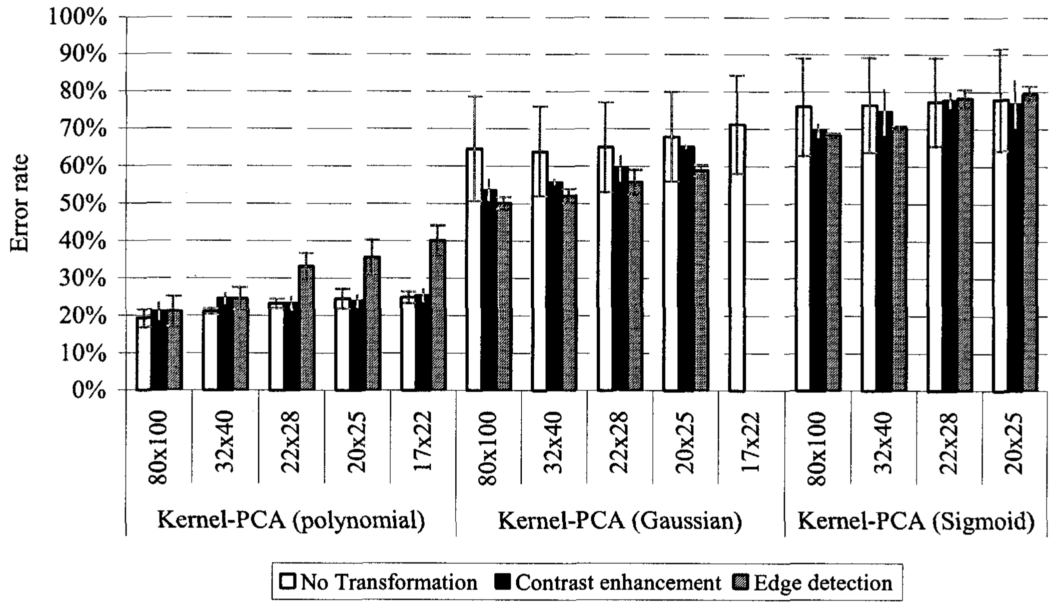


Figure 53 Performance of kernel-PCA using FERET