FUSION OF FACE AND GAIT FOR HUMAN RECOGNITION

by

RABIA JAFRI

(Under the Direction of Hamid R. Arabnia)

ABSTRACT

A system that integrates the face, a physical biometric, with gait, a behavioral biometric, for automatically recognizing human beings effectively under a wider range of conditions than a classifier which exclusively employs only one of these biometrics, is proposed. A decision-level fusion approach is adopted where the top matches of the face classifier are passed on to the gait classifier which then determines the identity of the unknown person. For face recognition, a principle components analysis-based approach, as well as a Bayesian inference-based classifier is employed, while for gait recognition, a model-based strategy is implemented, which utilizes various gait features identified as being the most pertinent for recognition based on data collected using an optoelectronic motion capture system. The integrated system is found to outperform the individual face and gait classifiers that it is composed of, thus, demonstrating the potential of using the gait to supplement the face in scenarios where the face classifier alone does not perform well due to the non-availability of high resolution face data.

During the course of this research, automated face recognition was also studied in detail, with a concentration on approaches that employ statistical dimensionality reduction techniques for this task. Experiments were conducted on some of the most widely-used methods in this category to test the recognition accuracy of these methods, and various combinations thereof, for different database sizes, images resolutions and number of bits per pixel. It was found that certain combinations of some of these techniques perform better than the individual methods that those combinations are comprised of when higher resolution face images are utilized. Furthermore, a system which utilizes different resolution versions of the whole face, as well as of various facial components, in a hierarchical manner was implemented and was found to achieve higher accuracy than its single-level counterpart which uses only the highest resolution images.

INDEX WORDS: Face recognition, Gait recognition, Integrated face-gait recognition, Fusion of biometrics, Person identification

FUSION OF FACE AND GAIT FOR HUMAN RECOGNITION

by

RABIA JAFRI

B.S., The University of Georgia, 1999

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

FUSION OF FACE AND GAIT FOR HUMAN RECOGNITION

by

RABIA JAFRI

Major Professor: Hamid R. Arabnia

Committee: Khaled Rasheed
Lynne Seymour
Thiab R. Taha

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

DEDICATION

To my parents, Syed Muhammad Nasim Jafri and Najma Jafri,

my husband, Syed Abid Ali,

and

my son, Omar.

ACKNOWLEDGEMENTS

I want to express my deep gratitude to my major professor, Dr. Hamid R. Arabnia, for his guidance and advice, unparalleled support and encouragement, and patience and understanding throughout my pursuit of this degree. Thank you for motivating me to stay in this program, for guiding me and helping me stay on track with your constant supervision, insightful comments and invaluable feedback, and for inspiring me to do my best. Without the advice, encouragement and opportunities that you provided, I would not have been able to accomplish this work.

I also want to extend my sincere thanks and appreciation to the members of my doctoral advisory committee, Dr. Khaled Rasheed, Dr. Lynne Seymour and Dr. Thiab R. Taha, for the great interest that they showed in my work and for their valuable advice and constructive suggestions.

I am profoundly grateful to Dr. Kathy J. Simpson for giving me the opportunity to learn about and use the optoelectronic motion capture system installed at the Biomechanics lab at the UGA. I sincerely appreciate all the assistance that she provided me in the process of gait data collection and analysis: assigning her graduate students to teach me how to operate the system and to aid me in collecting gait data, helping me recruit volunteers from her undergraduate classes for participation in gait data collection, and providing invaluable advice and recommendations regarding which gait parameters should be used for recognition. I also want to thank Dr. Simpson's graduate students, Sabrina Fordham and Yang-Chieh Fu, for assisting me in the gait data collection and the motion capture system set up.

My heartfelt gratitude goes to my family for their blessings, love and inspiration. I would obviously not be at this stage were it not for my parents' love and sacrifices. I am also thankful to my parents-in-law for supporting me financially and keeping me in their prayers. Most importantly, I would never have been able to come this far were it not for the love, support and encouragement that I received from my husband, Abid. Thanks for looking after our son while I attended classes and later, when I was teaching in the labs, for cheerfully improvising meals when I was too busy to cook, for discussing farfetched ideas for hours, for so patiently proofreading drafts of my papers and helping me to format them, and for coaxing me to continue with this research at times when I was ready to give it up.

Last but not least, I would like to thank my son, Omar (who turned six on the day I defended this dissertation), for putting up with a sleep-deprived, irritated mom one time too many and still responding with smiles and hugs. Thank you, little one, for all the joy you bring into my life!

Finally, I want to thank Allah for giving me the patience, strength, wisdom, support and opportunity to accomplish my goal.

Thank you all!

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1. Objective

Among the various biometric technologies explored in recent years, face recognition has emerged as one of the most promising biometric-based strategies to date with applications in domains such as surveillance, covert security and context-aware environments [37]. The face's appeal as a biometric stems from the several advantages that it offers in terms of being non-intrusive, non-invasive, cost-effective, easily accessible (i.e., face data can be conveniently acquired with a few inexpensive cameras) and relatively acceptable to the general public. Though current face recognition algorithms have reached a certain degree of maturity when operating under constrained conditions [37], however, there are some inherent weaknesses plaguing face recognition systems which diminish their effectiveness and essentially preclude their use in certain real-world situations (e.g., when the subject is too far away from the camera for his facial features to be clearly perceived, when the face is occluded or blurred or when the face is severely distorted due to extreme variations in expression, viewpoint, illumination, etc.) [38]. One obvious and feasible way to compensate for the degradation in the performance of the face recognition system in those scenarios is to use other biometric traits in conjunction with the face, which are not impeded by the same challenges as the face in those situations. Gait (a manner of walking or moving on foot) is a behavioral biometric characteristic which fits the bill.

The utility of gait for human recognition has only recently begun to be explored [39]. Similar to the face, gait, too, is a visual cue which can be extracted from video and thus, it appears to offer the same advantages (outlined above) that are presented by the face. However, gait offers some additional benefits in that, unlike the face, it can be acquired from a sequence of low resolution images of a person taken from a distance where the subject's body occupies too few pixels for other biometric traits to be discerned. Moreover, it is more robust to slight variations in viewpoint as compared to other biometrics and cannot be easily disguised without attracting attention. Nevertheless, gait as a biometric has its limitations since, being a behavioral trait, it can be affected and altered by factors such as clothing, footwear, environmental conditions, emotions, fatigue, drunkenness, pregnancy, injury, disease, aging and load.

The objective of this research is to supplement and combine the face, a physical biometric, with gait, a behavioral biometric, to construct a system capable of operating effectively under a wider range of conditions than a classifier which exclusively employs only one of these biometrics. As mentioned before, current face recognition techniques have been shown to perform reasonably well when operating on face data acquired under controlled conditions, and our experimentation with gait recognition has revealed that gait recognition systems also perform well when accurate data for a certain set of gait variables is available. Therefore, when presented with face and/or gait data for an unknown person whose identity is to be determined, the aim is to proceed as follows:

1) If high quality data for the face is available, recognition is performed based on the face alone.
2) In the absence of high quality data for the face, if high quality data for the gait is available, recognition is performed based on the gait alone.

3) If only low-quality data is available for both the face and the gait, then recognition is performed using a combination of the face with the gait.

1.2. Contributions

The main contribution of this research lies in the presentation of a novel system integrating the face with the gait and in the conclusions drawn from various tests used to evaluate this system.

During the course of this research, automated face recognition was also studied in detail and some methods were implemented in a bid to improve upon the performance of existing subspace-based techniques for this task. The results obtained from these endeavors constitute some secondary contributions.

The following discussion summarizes the specific contributions made in the areas mentioned above.

1.2.1. Integration of the face with the gait

For fusing the face and gait, a decision-level fusion approach is adopted where the top matches of the face classifier are passed on to the gait classifier which then determines the identity of the unknown person. For face recognition, the eigenfaces approach [6], as well as a Bayesian inference-based classifier from our previous work (described in [40]) is employed, while for gait recognition, a model-based (rather than holistic) strategy is implemented, which utilizes various gait features identified as being the most pertinent for recognition based on data collected using an optoelectronic motion capture system.

The above system differs from previously presented integrated face-gait techniques in the following ways (Note that only a few approaches for integrating the face and the gait have been proposed in the literature [41]): Previous methods for fusing the face and gait have opted to employ silhouette-based holistic strategies for recognizing the gait and have not incorporated more detailed gait features in their characterization of the gait. Our system, on the other hand, employs a model-based strategy that utilizes dynamic gait features, i.e., spatio-temporal variables (e.g., walking speed and stride length) and joint angle trajectories (e.g, elbow, knee and ankle angles), to form the gait signature. Moreover, though the eigenfaces method has been used for face classification in previously proposed face-gait recognition systems, however, a Bayesian-inference based face classifier has not been utilized until now in such an approach.

Our gait recognition strategy is also different from previous model-based techniques as follows: Though several model-based techniques for gait recognition have previously been proposed [42-44], however, these strategies have utilized only a limited set of gait features extracted from video sequences. Our gait recognition method differs from previous approaches in that it utilizes a much more extensive set of gait features and in that those features are extracted from motion capture data rather than from video. The decision to use motion capture data was motivated by two considerations: 1) the need to avoid inaccuracies in gait data stemming from the usual problems of extracting visual cues from video such as imperfect foreground segmentation of the walking subject from the background scene and poor imaging conditions. 2) to facilitate the extraction of new variables that have not previously been utilized for recognizing the gait and to determine if there is any useful identity information present in those features before even attempting to recover those features from a video sequence.

Some of the findings from the empirical evaluation of this system are outlined below:

- Both spatio-temporal variables and angle trajectories were found to be useful for recognition, though the angle trajectories yielded higher recognition rates as compared to the spatio-temporal variables.
- Step time and cadence emerged as the best-performing spatio-temporal variables while the hip angle trajectories emerged as the best-performing angle trajectories.
- The lower body angle trajectories (i.e., those of the ankles, knees and hips, foot progression and pelvis angles) were found to perform better than the upper body ones.
- When angle trajectories that produced low recognition rates individually were combined, the resulting combinations produced much higher recognition rates than the individual angles.
- The integrated face-gait classifiers outperformed the individual face and gait classifiers that they were composed of for several different sets of gait variables, thus, demonstrating the potential of using the gait to supplement the face in scenarios where the face classifier alone does not perform well due to the non-availability of high resolution face data.
- These experiments were repeated using a larger data set consisting of the 24 original subjects and 132 synthetic subjects (whose gaits were generated from the 24 real subjects). A comparison of the results obtained with this larger data set with those previously obtained with the smaller data set showed that though most of the previous observations and conclusions remained true, however, the utility of certain gait variables, such as the spatio-temporal variables, for recognition decreased as the number of subjects was increased. However, in general, the integrated face-gait system did continue to

outperform the individual face and gait classifiers that it was composed of, for several different sets of gait variables, demonstrating the viability of this approach even for larger data sets.

1.2.2. Improving subspace-based face recognition methods

The focus of the early part of this doctoral research has been on automated face recognition, with a concentration on approaches that employ statistical dimensionality reduction techniques for this task. The experiments that were conducted on some of the most widely-used methods in this category (i.e., the PCA-based Eigenfaces method, the LDA-based Fisherfaces technique and the Bayesian-inference-based probabilistic Eigenfaces approach) to test the recognition accuracy of these methods, and various combinations thereof, for different database sizes, images resolutions and number of bits per pixel, have revealed the following novel results, among others:

- The recognition accuracy of the Fisherfaces method deteriorates drastically relative to the other methods with a decrease in resolution. The recognition performance of this technique is also significantly decreased when the number of bits per pixel is reduced (the other methods are relatively unaffected by the change in the number of bits per pixel).
- Certain combinations of these techniques perform better than the individual methods that those combinations are comprised of at higher resolutions (the improvement in recognition accuracy is up to 4 percentage points which is quite significant).

At present, most face recognition methods utilize the image of the whole face to perform classification. However, in recent years, several techniques have been proposed which segment the face image into various parts, apply known face recognition methods to those facial regions and finally, combine the recognition results thus obtained to determine the identity of the whole face. These methods have been reported to produce significantly better recognition rates as compared to their holistic counterparts under variations in expression, illumination, decoration, head rotation and short aging [45]. The encouraging results for these component-based techniques obtained from my own experimentation as well as those reported by others motivated me to further investigate some variations of these approaches which involve the concept of hierarchy. One such system, which has yielded good results in terms of recognition accuracy, utilizes different resolution versions of the whole face, as well as various facial components. Each level of the hierarchy uses different components and higher resolution images as compared to the previous layer and reduces the size of the gallery to be examined by the next layer. The idea here is to filter out the most similar faces at the first level using a minimum amount of information per image and then to feed those faces to subsequent levels each of which utilizes more information per image and returns a smaller (and hopefully, more accurate) set of the most similar images than the previous level. Experiments conducted to test the performance of the proposed system using frontal images from the FERET database and a 2D-PCA based face recognition method indicate that this technique achieves higher accuracy than its single-level counterpart which utilizes the highest resolution images.

1.3. Organization of the dissertation

An overview of the chapters in this dissertation and how they relate to each other is provided below:

Chapter 1 defines the objectives of this research and delineates its contributions. It also describes how this dissertation is organized.

Chapter 2 consists of a survey of face recognition techniques. In this chapter, an overview of some of the well-known methods in various categories of face recognition techniques is provided and some of the benefits and drawbacks of the schemes mentioned therein are examined. Furthermore, a discussion outlining the incentive for using face recognition, the applications of this technology and some of the difficulties plaguing current systems for this task has also been provided. This chapter also mentions some of the most recent algorithms developed for this purpose and attempts to give an idea of the state of the art in face recognition technology.

In Chapter 3, the recognition accuracy of three face recognition algorithms based on subspace methods (i.e., the Eigenfaces approach, the Fisherfaces technique, and the Bayesian inference-based probabilistic Eigenfaces approach) is tested for different gallery sizes, image resolutions and number of bits per pixel. The same experiments are then repeated for various combinations of these techniques and the performance of these combinations with that of the individual methods is compared.

Based on the favorable results reported in face recognition literature for methods utilizing the statistical technique of Principal Components Analysis (PCA), as well as the encouraging results obtained in my own experimentation with this technique, a study of the different approaches based on this method, with a focus on those which differ from the classical PCA-based face recognition method only in terms of the face image representation being used, was conducted and is presented in Chapter 4.

At present, most face recognition methods utilize the image of the whole face to perform classification. However, in recent years, several techniques have been proposed which segment the face image into various parts, apply known face recognition methods to those facial regions and finally, combine the recognition results thus obtained to determine the identity of the whole face. A survey of some of the most widely known methods in this category is presented in Chapter 5 and a detailed analysis of their performance as well as a discussion of how they compare to approaches which utilize the whole face without subdividing it is provided.

The encouraging results reported for PCA-based and component-based face recognition techniques (described in Chapters 4 and 5) motivated me to further investigate some variations of these approaches. One such method which involves the concept of hierarchy is presented in Chapter 6. This approach utilizes different resolution versions of the whole face, as well as various facial components. Each level of the hierarchy uses different components and higher resolution images as compared to the previous layer and reduces the size of the gallery to be examined by the next layer. The idea here is to filter out the most similar faces at the first level using a minimum amount of information per image and then to feed those faces to subsequent levels each of which utilizes more information per image and returns a smaller (and hopefully, more accurate) set of the most similar images than the previous level.

A variation of the above component-based strategy is introduced in Chapter 7, where the recognition process is again divided into several levels, each level utilizing a less discriminatory subset of facial components as compared to the previous level in order to reduce the set of images to be examined at the next level. The objective is to use a relatively simple classifier to filter out the most similar faces at the first level and then to feed those faces to classifiers which look at the finer details at subsequent levels. Moreover, this appears to be in line with the way that human beings recognize faces [46, 47].

As mentioned before, current subspace-based face recognition techniques have been shown to perform reasonably well when operating under controlled conditions, and the work presented in chapters 2 and 6 shows that unifying these approaches and modifying them has the potential of further improving their performance. Furthermore, the results presented in chapter 2 as well as other studies [48], have demonstrated that the performance of some well-known subspace-based face recognition algorithms remains fairly stable with decreasing image resolution, with good recognition rates being achieved for resolutions as low as 64x64 pixels. However, these studies have also shown that when the image resolution is decreased even further, the performance of these algorithms degrades dramatically. Given the robustness of face recognition methods for resolutions as low as 64x64 pixels, recognition based on the face is a viable option even for applications using low-resolution face imagery. However, some mechanism is needed to compensate for the degradation in performance experienced by the face classifier when the image resolution falls below this threshold. One feasible way to achieve this is to use some other biometric in conjunction with the face which is independent of the face but offers the same advantages in terms of being non-intrusive and non-invasive. Gait (a manner of walking or moving on foot) is a biometric which meets these requirements.

Some techniques have already been proposed in recent years which integrate face, a physical biometric, with gait, a behavioral biometric, with the aim of investigating if such a combination will improve upon the performance of methods which exclusively employ only one of these biometrics. An overview of some of the well-known approaches in this area, along with a discussion of the advantages offered and the challenges faced by such systems, is provided in Chapter 8 and the potential of this technology for further research and application is explored.

A system that integrates the face, a physical biometric, with gait, a behavioral biometric, for automatically recognizing human beings is presented in chapter 9. A decision-level fusion approach is adopted where the top matches of the face classifier are passed on to the gait classifier which then determines the identity of the unknown person. For gait recognition, a novel system is implemented, which utilizes various gait features identified as being the most pertinent for recognition based on data collected using an optoelectronic motion capture system.

The integrated face-gait approach presented in chapter 9 is evaluated using a larger data set consisting of the 24 original subjects and 132 synthetic subjects (whose gaits were generated from the 24 real subjects) and the results are reported in chapter 10.

Conclusions and proposals for future work are presented in chapter 11.

CHAPTER 2

A SURVEY OF FACE RECOGNITION TECHNIQUES

2.1. Problem definition

The face recognition problem can be formulated as follows: Given an input face image and a database of face images of known individuals, verify or determine the identity of the person in the input image.

2.2. Why use the face for recognition

Biometric-based techniques have emerged as the most promising option for recognizing individuals in recent years since, instead of authenticating people and granting them access to physical and virtual domains based on passwords, PINs, smart cards, plastic cards, tokens, keys etc., these methods examine an individual's physical and/or behavioral characteristics in order to determine and/or ascertain his identity. Passwords and PINs are hard to remember and can be stolen or guessed; cards, tokens, keys etc. can be misplaced, forgotten, purloined or duplicated; magnetic cards can become corrupted and unreadable. However, an individual's biological traits cannot be misplaced, forgotten, stolen or forged.

Biometric-based technologies include identification based on physiological characteristics (such as face, fingerprints, finger geometry, hand geometry, hand veins, palm, iris, retina, ear and voice) and behavioral traits (such as gait, signature and keystroke dynamics)

[49]. Face recognition appears to offer several advantages over other biometric methods, a few of which are outlined here: almost all these technologies require some voluntary action by the user, i.e., the user needs to place his hand on a hand rest for fingerprinting or hand geometry detection and has to stand in a fixed position in front of a camera for iris or retina identification. However, face recognition can be done passively without any explicit action or participation from the user since face images can be acquired from a distance by a camera. This is particularly beneficial for security and surveillance purposes. Furthermore, data acquisition in general is fraught with problems for other biometrics: techniques that rely on hands and fingers can be rendered useless if the epidermis tissue is damaged in some way (i.e., bruised or cracked). Iris and retina identification require expensive equipment and are much too sensitive to any body motion. Voice recognition is susceptible to background noises in public places and auditory fluctuations on a phone line or tape recording. Signatures can be modified or forged. However, facial images can be easily obtained with a couple of inexpensive fixed cameras. Good face recognition algorithms and appropriate preprocessing of the images can compensate for noise and slight variations in orientation, scale and illumination. Finally, technologies that require multiple individuals to use the same equipment to capture their biological characteristics potentially expose the user to the transmission of germs and impurities from other users. However, face recognition is totally non-intrusive and does not carry any such health risks.

2.3. Applications

Face recognition is used for two primary tasks:

1. Verification (one-to-one matching): When presented with a face image of an unknown individual along with a claim of identity, ascertaining whether the individual is who he/she claims to be.

2. Identification (one-to-many matching): Given an image of an unknown individual, determining that person's identity by comparing (possibly after encoding) that image with a database of (possibly encoded) images of known individuals.

There are numerous application areas in which face recognition can be exploited for these two purposes, a few of which are outlined below.

- Security (access control to buildings, airports/seaports, ATM machines and border checkpoints [50, 51]; computer/network security [52]; email authentication on multimedia workstations).
- Surveillance (a large number of CCTVs can be monitored to look for known criminals, drug offenders etc. and authorities can be notified when one is located; for example, this is what was done at the Super Bowl 2001 game at Tampa, Florida [53]; in another instance, according to a CNN report, two cameras linked to state and national databases of sex offenders, missing children and alleged abductors have been installed recently at Royal Palm Middle School in Phoenix, Arizona [54]).
- General identity verification (electors registration, banking, electronic commerce, identifying newborns, national IDs, passports, drivers licenses, employee IDs).
- Criminal justice systems (mug-shot/booking systems, post-event analysis, forensics).
- Image database investigations (searching image databases of licensed drivers, benefit recipients, missing children, immigrants and police bookings).
- "Smart Card" applications (in lieu of maintaining a database of facial images, the faceprint can be stored in a smart card, bar code or magnetic stripe, and authentication is performed by matching the live image and the stored template) [32].

- Multi-media environments with adaptive human-computer interfaces (part of ubiquitous or context-aware systems, behavior monitoring at childcare or old people center, recognizing a customer and assessing his needs) [55, 56].
- Video indexing (labeling faces in video) [57, 58].
- Witness face reconstruction [59].

In addition to these applications, the underlying techniques in the current face recognition technology have also been modified and used for related applications such as gender classification [60-62], expression recognition [63, 64] and facial feature recognition and tracking [65]; each of these has its utility in various domains: for instance, expression recognition can be utilized in the field of medicine for intensive care monitoring [66] while facial feature recognition and detection can be exploited for tracking a vehicle driver's eyes and thus, monitoring his fatigue [67] as well as for stress detection [68].

Face recognition is also being used in conjunction with other biometrics such as speech, iris, fingerprint, ear and gait recognition in order to enhance the recognition performance of these methods [34, 55, 69-80].

2.4. General difficulties

Face recognition is a specific and hard case of object recognition. The difficulty of this problem stems from the fact that in their most common form (i.e., the frontal view) faces appear to be roughly alike and the differences among them are quite subtle. Consequently, frontal face images form a very dense cluster in image space which makes it virtually impossible for traditional pattern recognition techniques to accurately discriminate among them with a high degree of success [81].

Furthermore, the human face is not a unique rigid object. There are numerous factors that cause the appearance of the face to vary. The sources of variation in the facial appearance can be categorized into two groups: intrinsic factors and extrinsic ones [82]. A) Intrinsic factors are due purely to the physical nature of the face and are independent of the observer. These factors can be further divided into two classes: intrapersonal and interpersonal [83]. Intrapersonal factors are responsible for varying the facial appearance of the same person, some examples being age, facial expression and facial paraphernalia (facial hair, glasses, cosmetics etc.). Interpersonal factors, however, are responsible for the differences in the facial appearance of different people, some examples being ethnicity and gender. B) Extrinsic factors cause the appearance of the face to alter via the interaction of light with the face and the observer. These factors include illumination, pose, scale and imaging parameters (e.g., resolution, focus, imaging noise etc.).

Evaluations of state of the art recognition techniques held during the past several years, such as the FERET evaluations [32, 84], FRVT 2000 [85], FRVT 2002 [86] and the FAT 2004 [87], have confirmed that age variations, illumination variations and pose variations are three major problems plaguing current face recognition systems [88].

Although most current face recognition systems work well under constrained conditions (i.e., scenarios in which at least a few of the factors contributing to the variability among face images are controlled), however, the performance of most of these systems degrades rapidly when they are put to work under conditions where none of these factors are regulated [89].

2.5. Face recognition techniques

The method for acquiring face images depends upon the underlying application. For instance, surveillance applications may best be served by capturing face images by means of a video camera while image database investigations may require static intensity images taken by a standard camera. Some other applications, such as access to top security domains, may even necessitate the forgoing of the non-intrusive quality of face recognition by requiring the user to stand in front of a 3D scanner or an infra-red sensor. Therefore, depending on the face data acquisition methodology, face recognition techniques can be broadly divided into three categories: methods that operate on intensity images, those that deal with video sequences, and those that require other sensory data such as 3D information or infra-red imagery. The following discussion sheds some light on the methods in each category and attempts to give an idea of some of the benefits and drawbacks of the schemes mentioned therein in general (for detailed surveys, please see [37, 90]).

2.5.1. Face recognition from intensity images

Face recognition methods for intensity images fall into two main categories: feature-based and holistic [2, 91, 92]. An overview of some of the well-known methods in these categories is given below.

2.5.1.1. Feature-based. Feature-based approaches first process the input image to identify and extract (and measure) distinctive facial features such as the eyes, mouth, nose, etc. as well as other fiducial marks and then compute the geometric relationships among those facial points, thus reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed for matching faces using these measurements.

Early work done on automated face recognition was mostly based on these techniques. One of the earliest of such attempts was by Kanade [93] who used simple image processing methods to extract a vector of 16 facial parameters which were ratios of distances, areas and angles (to compensate for the varying size of the pictures) and used a simple Euclidean distance measure for matching to achieve a peak performance of 75% on a database of 20 different people using 2 images per person (one for reference and one for testing).

Brunelli and Poggio [2] built upon Kanade's approach and computed a vector of 35 geometric features (Figure 2.1) from a database of 47 people (4 images per person) and reported a 90% recognition rate. However, they also reported 100% recognition accuracy for the same database using a simple template-matching approach.

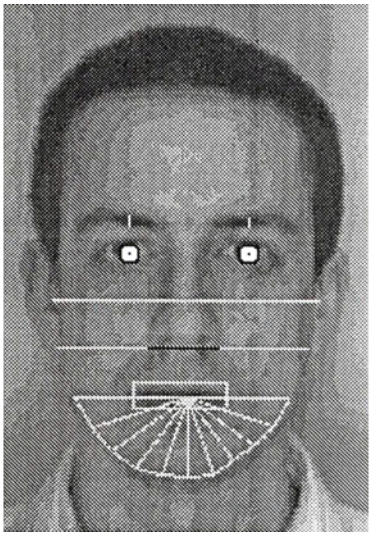


Figure 2.1. Geometrical features (white) used in the face recognition experiments [2].

(©1993 IEEE)

More sophisticated feature extraction techniques involve deformable templates ([94], [95], [96]), Hough transform methods [97], Reisfeld's symmetry operator [98] and Graf's filtering and morphological operations [99]. However, all of these techniques rely heavily on heuristics such as restricting the search subspace with geometrical constraints [100]).

Furthermore, a certain tolerance must be given to the models since they can never perfectly fit the structures in the image. However, the use of a large tolerance value tends to destroy the precision required to recognize individuals on the basis of the model's final best-fit parameters and makes these techniques insensitive to the minute variations needed for recognition [83]. More recently, Cox et al. [101] reported a recognition performance of 95% on a database of 685 images (a single image for each individual) using a 30-dimensional feature vector derived from 35 facial features (Figure 2.2). However, the facial features were manually extracted, so it is reasonable to assume that the recognition performance would have been much lower if an automated and, hence, less precise, feature extraction method had been adopted. In general, current algorithms for automatic feature extraction do not provide a high degree of accuracy and require considerable computational capacity [10].

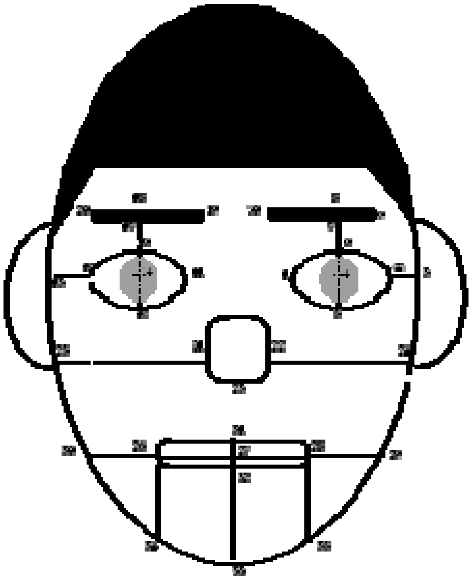


Figure 2.2. 35 manually identified facial features [101]. (©1996 IEEE)

Another well-known feature-based approach is the elastic bunch graph matching method proposed by Wiskott et al. [102] . This technique is based on Dynamic Link Structures [103]. A graph for an individual face is generated as follows: a set of fiducial points on the face are chosen. Each fiducial point is a node of a full connected graph, and it is labeled with the Gabor filters responses applied to a window around the fiducial point. Each arch is labeled with the distance between the correspondent fiducial points. A representative set of such graphs is combined into a stack-like structure, called a *face bunch graph*. Once the system has a face bunch graph, graphs for new face images can then be generated automatically by Elastic Bunch Graph Matching. Recognition of a new face image is performed by comparing its image graph to those of all the known face images and picking the one with the highest similarity value. Using this architecture, the recognition rate can reach 98% for first rank and 99% for the first 10 ranks using a gallery of 250 individuals. The system has been enhanced to allow it to deal with different poses (Figure 2.3) [104] but the recognition performance on faces of the same orientation remains the same. Though this method was among the best performing ones in the most recent FERET evaluation [105, 106], however, it suffers from the serious drawback of requiring the graph placement for the first 70 faces to be done manually before the elastic graph matching becomes adequately dependable [107]. Campadelli and Lanzarotti [108] have recently experimented with this technique where they have eliminated the need to do the graph placement manually by using parametric models, based on the deformable templates proposed in [94], to automatically locate fiducial points. They claim to have obtained the same performances as the elastic bunch graph employed in [102]. Other recent variations of this approach replace the Gabor features by a graph matching strategy [109] and HOGs (Histograms of Oriented Gradients [110].

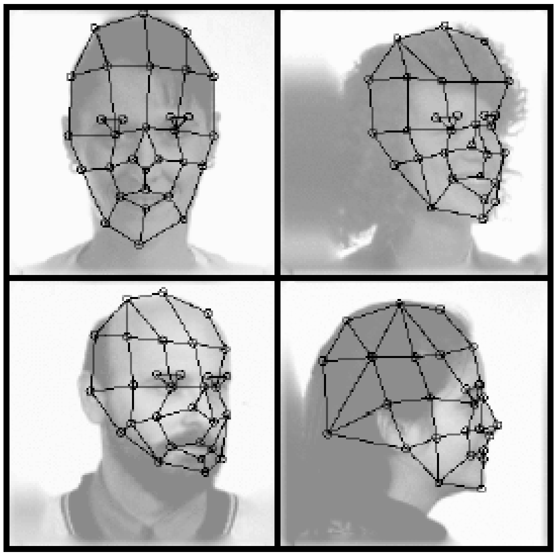


Figure 2.3. Grids for face recognition [104]. (©1999 IEEE)

Considerable effort has also been devoted to recognizing faces from their profiles [111-115] since, in this case, feature extraction becomes a somewhat simpler one-dimensional problem [101, 114]. Kaufman and Breeding [113] reported a recognition rate of 90% using face profiles; however, they used a database of only 10 individuals. Harmon et al. [111] obtained recognition accuracies of 96% on a database of 112 individuals using a 17-dimensional feature vector to describe face profiles and utilizing a Euclidean distance measure for matching. More recently, Liposcak and Loncaric [114] reported a 90% accuracy rate on a database of 30 individuals using subspace filtering to derive a 21-dimensional feature vector to describe the face profiles and employing a Euclidean distance measure to match them (Figure 2.4).

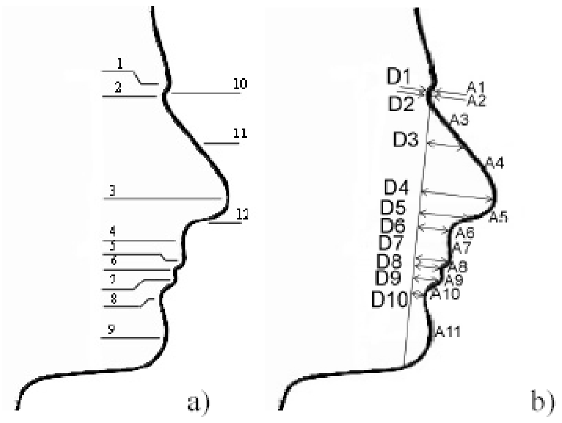


Figure 2.4. a) The twelve fiducial points of interest for face recognition, b) Feature vector has 21 components; ten distances D_1-D_{10} (normalized with /(D_4+D_5)) and eleven profile arcs A_1-A_{11} (normalized with /(A_5+A_6)) [114]. (Courtesy of Z. Liposcak and S. Loncaric)

2.5.1.1.1. Advantages and disadvantages. The main advantage offered by the featured-based techniques is that since the extraction of the feature points precedes the analysis done for matching the image to that of a known individual, such methods are relatively robust to position variations in the input image [83]. In principle, feature-based schemes can be made invariant to size, orientation and/or lighting [101]. Other benefits of these schemes include compactness of representation of the face images and high speed matching [116].

The major disadvantage of these approaches is the difficulty of automatic feature detection (as discussed above) and the fact that the implementer of any of these techniques has to make arbitrary decisions about which features are important [117]. After all, if the feature set lacks discrimination ability, no amount of subsequent processing can compensate for that intrinsic deficiency [101].

2.5.1.2. Holistic. Holistic approaches attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. These schemes can be subdivided into two groups: statistical and AI approaches. An overview of some of the methods in these categories follows.

2.5.1.2.1. Statistical. In the simplest version of the holistic approaches, the image is represented as a 2D array of intensity values and recognition is performed by direct correlation comparisons between the input face and all other faces in the database. Though this approach has been shown to work [118] under limited circumstances (i.e., equal illumination, scale, pose etc.), it is computationally very expensive and suffers from the usual shortcomings of straightforward correlation-based approaches, such as sensitivity to face orientation, size, variable lighting conditions, background clutter and noise [119]. The major hindrance to the direct-matching methods' recognition performance is that they attempt to perform classification in a space of very high dimensionality [119]. To counter this curse of dimensionality, several other schemes have been proposed that employ statistical dimensionality reduction methods to obtain and retain the most meaningful feature dimensions before performing recognition. A few of these are mentioned below.

Sirovich and Kirby [120] were the first to utilize Principal Components Analysis (PCA) [121, 122] to economically represent face images. They demonstrated that any particular face can be efficiently represented along the eigenpictures coordinate space, and that any face can be approximately reconstructed by using just a small collection of eigenpictures and the corresponding projections ('coefficients') along each eigenpicture.

Turk and Pentland [6, 123] realized, based on Sirovich and Kirby's findings, that projections along eigenpictures could be used as classification features to recognize faces. They employed this reasoning to develop a face recognition system that builds eigenfaces, which correspond to the eigenvectors associated with the dominant eigenvalues of the known face (patterns) covariance matrix, and then recognizes particular faces by comparing their projections along the eigenfaces to those of the face images of the known individuals. The eigenfaces define a feature space that drastically reduces the dimensionality of the original space, and face identification is carried out in this reduced space. An example training set, the average face and the top seven eigenfaces derived from the training images are shown in Figures 2.5, 2.6 and 2.7,

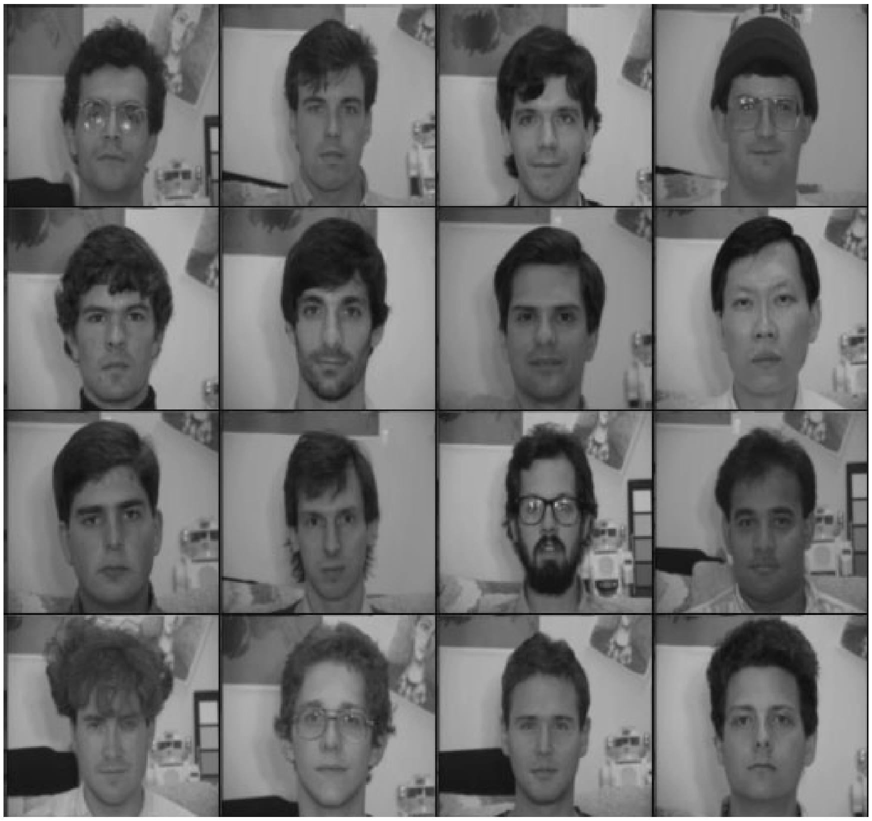


Figure 2.5. An example training set [6]. (With kind permission of MIT Press Journals)

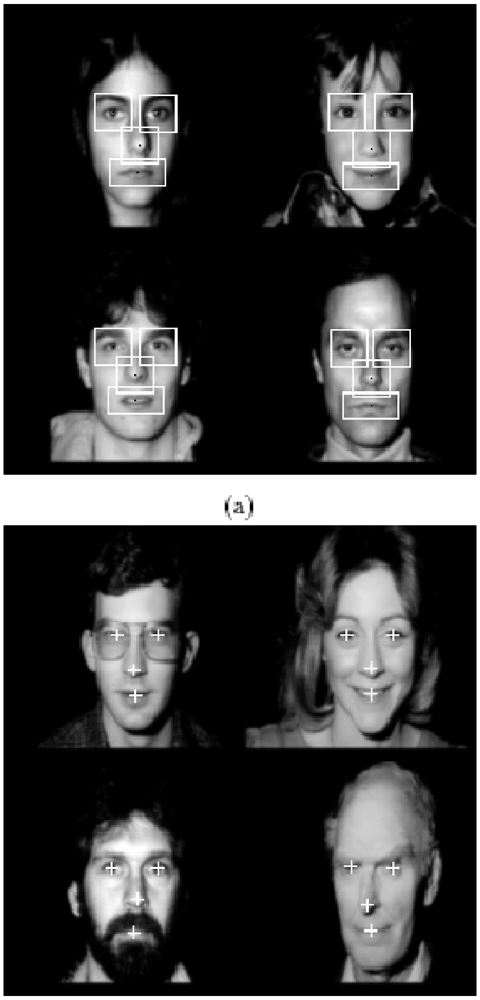
Figure 2.6. The average face [6]. (With kind permission of MIT Press Journals)

Figure 2.7. Seven of the eigenfaces calculated from the images of Figure 2.5, without the background removed [6]. (With kind permission of MIT Press Journals)

respectively. The method was tested using a database of 2500 images of 16 people under all combinations of 3 head orientations, 3 head sizes or scales and 3 lighting conditions and various resolutions. Recognition rates of 96%, 85% and 64% were reported for lighting, orientation and scale variation. Though the method appears to be fairly robust to lighting conditions, but its performance degrades with scale changes.

The capabilities of Turk and Pentland's system have been extended in several ways in [1] and tested on a database of 7562 images of approximately 3000 people. A "multiple observer" method has been suggested to deal with large changes in pose: Given N individuals under M different views, one can either do recognition and pose estimation in a universal eigenspace calculated from the combination of NM images (parametric approach) or alternatively, one can build a set of M separate eigenspaces, one for each of the N views (the view-based approach). The view-based approach is reported to have yielded better results than the parametric one. A modular "eigenfeatures" approach has also been proposed to deal with localized variations in the face appearance where a low-resolution description of the whole face is augmented by additional higher resolution details in terms of salient facial features (Figure 2.8). This system is reported to have produced slightly better results than the basic eigenfaces approach (Figures 2.9, 2.10). Though no implementation is reported, however, it has been suggested in [6] that variation in scale be dealt with by employing multiscale eigenfaces or by rescaling the input image to multiple sizes and using the scale that results in the smallest distance measure to the face space.

Though PCA appears to work well when a single image of each individual is available, however, when multiple images per person are present, then Belhumeur et al. [8] argue that by choosing the projection which maximizes total scatter, PCA retains unwanted variations due to lighting and facial expression. As stated by Moses et al. [124], "the variations between the images of the same face due to illumination and lighting direction are almost always larger than image variations due to change in face identity" (see Figure 2.11 for an example of this.) They, therefore, propose using Fisher's Linear Discriminant Analysis [125] which maximizes the ratio of the between-class scatter and the within-class scatter and thus, is purportedly better for



(a)

(b)

Figure 2.8. (a) Examples of facial feature training templates used and (b) the resulting typical detections [1]. (©1994 IEEE)

classification than PCA. Conducting various tests on 330 images of 5 people (66 of each), they report that their method, called Fisherfaces, which uses subspace projection prior to LDA projection (to prevent the within-class scatter matrix from becoming degenerate), is better at simultaneously handling variation in lighting and expression. Swets and Weng [9] previously reported similar results employing the same procedure not only for faces but also for general

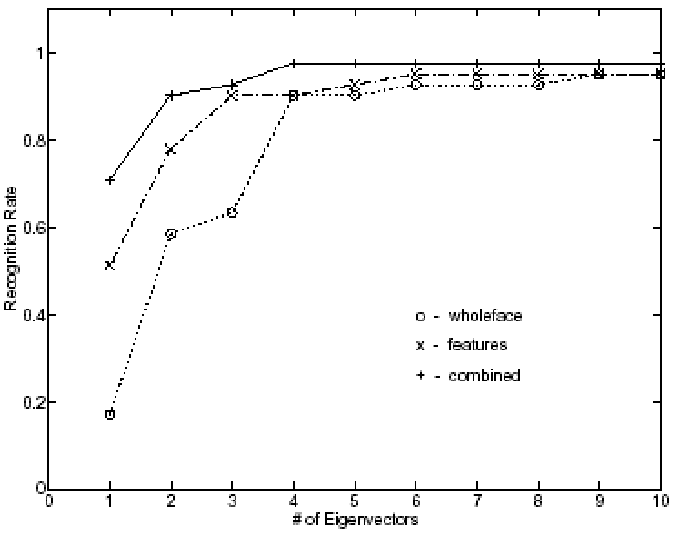


Figure 2.9. Recognition rates for eigenfaces, eigenfeatures and the combined modular representation [1]. (©1994 IEEE)

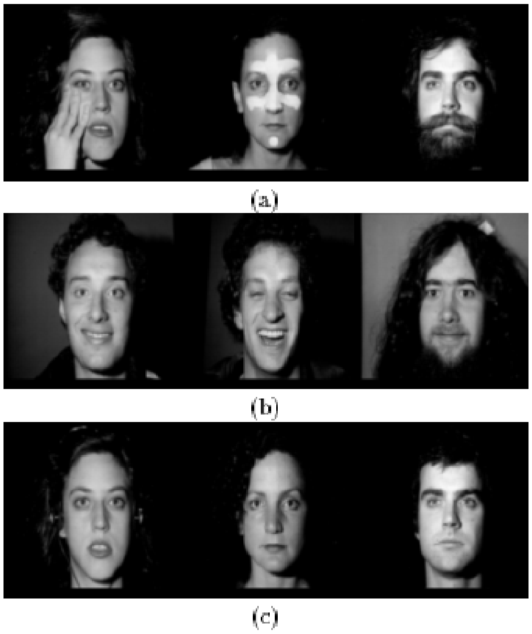


Figure 2.10. (a) Test views, (b) Eigenface matches, (c) Eigenfeature matches [1]. (©1994 IEEE)

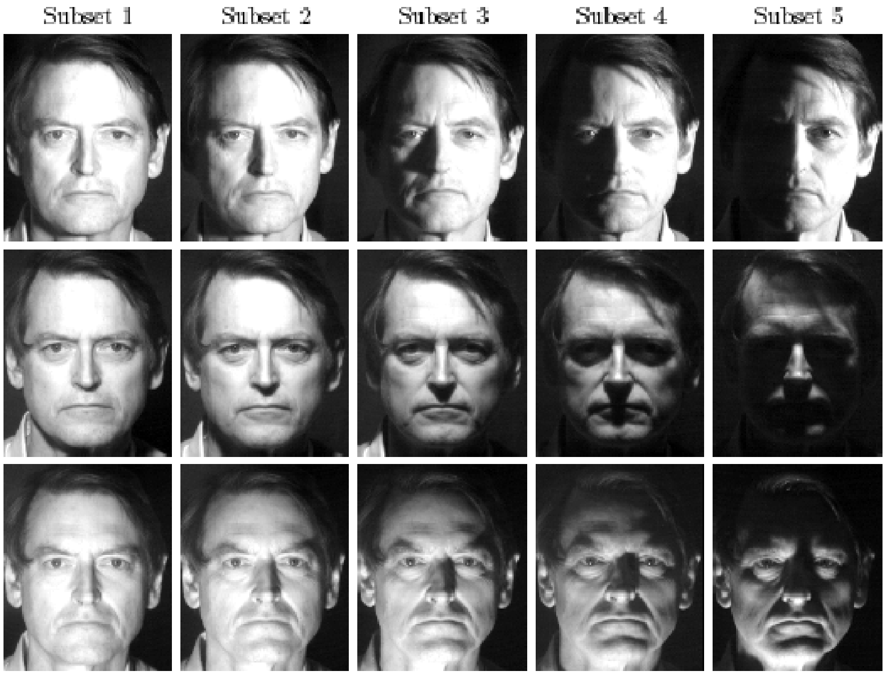


Figure 2.11. The same person seen under varying light conditions can appear dramatically different [8]. (©1997 IEEE)

objects (90% accuracy on a database of 1316+298 images from 504 classes) (Figure 2.12 shows some examples of eigenfaces and Fisherfaces and how Fisherfaces capture discriminatory information better than eigenfaces). It should be noted, however, that some recent work [126] shows that when the training data set is small, PCA can outperform LDA and also that PCA is less sensitive to different training sets.

The standard eigenfaces and the Fisherfaces approaches assume that an optimal projection exists that projects the face images to distinct non-overlapping regions in the reduced subspace where each of these regions corresponds to a unique subject. However, in reality, that assumption may not necessarily be true since images of different people may frequently map to the same region in the face space and, thus, the regions corresponding to different individuals may not always be disjoint.

(a)

(b)

Figure 2.12. (a) This sample of eigenfaces shows the tendency of principal components to capture major variations in the training set such as lighting direction; (b) The corresponding sample of Fisherfaces shows the ability of Fisherfaces to discount those factors unrelated to classification [9]. (©1996 IEEE)

Moghaddam et al. [127] propose an alternative approach which utilizes difference images, where a difference image for two face images is defined as the signed arithmetic difference in the intensity values of the corresponding pixels in those images. Two classes of difference images are defined: *intrapersonal*, which consists of difference images originating from two images of the same person, and *extrapersonal,* which consists of difference images derived from two images of different people.

It is assumed that both these classes originate from discrete Gaussian distributions within the space of all possible difference images. Then, given the difference image between two images I_1 and I_2, the probability that the difference image belongs to the intrapersonal class is given by Bayes Rule as follows:

$$P(\Omega_I \mid d(I_1, I_2)) = \frac{P(d(I_1, I_2) \mid \Omega_I) P(\Omega_I)}{P(d(I_1, I_2) \mid \Omega_I) P(\Omega_I) + P(d(I_1, I_2) \mid \Omega_E) P(\Omega_E)} \qquad \textbf{(2.1)}$$

where

$d(I_1, I_2)$ = the difference image between two images I_1 and I_2

Ω_I = the intrapersonal class

Ω_E= the extrapersonal class

The formulation of the face recognition problem in this manner converts it from an *m*-ary classification problem (where *m* is the number of individuals in the database of known face images) into a binary classification problem which can be solved using the maximum a posteriori (MAP) rule – the two images are declared to belong to the same individual if $P(\Omega_I|d(I_1, I_2)) > P(\Omega_E|d(I_1, I_2))$ or equivalently, if $P(\Omega_I|d(I_1, I_2)) > ½$. For a computationally more expedient approach, Moghaddam and Pentland [128] also suggest ignoring the extrapersonal class information and calculating the similarity based only on the intrapersonal class information. In the resulting maximum likelihood (ML) classifier, the similarity score is given only by $P(d(I_1, I_2)| \Omega_I)$.

Numerous variations/extensions to the standard eigenfaces and the Fisherfaces approaches have been suggested since their introduction. Some recent advances in PCA-based algorithms include multi-linear subspace analysis [129], symmetrical PCA [130], two-dimensional PCA [131, 132], eigenbands [133], adaptively weighted sub-pattern PCA [134], weighted modular PCA [26], Kernel PCA [135, 136] and diagonal PCA [21]. Examples of recent LDA-based algorithms include Direct LDA [137, 138], Direct-weighted LDA [139], Nullspace LDA [140, 141], Dual-space LDA [142], Pair-wise LDA [143], Regularized Discriminant Analysis [144], Generalized Singular Value Decomposition [145, 146], Direct Fractional-Step LDA [147], Boosting LDA [148], Discriminant Local Feature Analysis [149], Kernel PCA/LDA [150, 151], Kernel Scatter-Difference Based Discriminant Analysis [152], 2D-LDA [153, 154], Fourier-LDA [155], Gabor-LDA [156], Block LDA [157], Enhanced FLD [158], Component-based Cascade LDA [159], and incremental LDA [160], to name just a few. All these methods purportedly obtain better recognition results than the baseline techniques.

One main drawback of the PCA and LDA methods is that these techniques effectively see only the Euclidean structure and fail to discover the underlying structure if face images lie on a non-linear submanifold in the image space. Since it has been shown that face images possibly reside on a nonlinear submanifold [161-167] (especially if there is a perceivable variation in viewpoint, illumination or facial expression), therefore, some nonlinear techniques have also been proposed to discover the nonlinear structure of the manifold, e.g., Isometric Feature Mapping (ISOMAP) [167], Locally Linear Embedding (LLE) [163, 168], Laplacian Eigenmap [169], Locality Preserving Projection (LPP) [170], Embedded Manifold [171], Nearest Manifold Approach [172], Discriminant Manifold Learning [173] and Laplacianfaces [174].

The eigenvectors found by PCA depend only on pair wise relationships between pixels in the image database. However, other methods exist that can find basis vectors that depend on higher-order relationships among the pixels and it seems reasonable to expect that utilizing such techniques would yield yet better recognition results. Independent component analysis (ICA) [175], a generalization of PCA, is one such method that has been employed for the face recognition task. ICA aims to find an independent, rather than an uncorrelated image decomposition and representation. Bartlett et al. [176] performed ICA on images in the FERET database under two different architectures, one which treated the images as random variables and the pixels as outcomes, and a second which treated the pixels as the random variables and the images as outcomes. Both ICA representations outperformed PCA representations for recognizing faces across days and changes in expression. A classifier that combined both ICA representations gave the best performance. Others have also experimented with ICA [177-183] and have reported that this technique, and variations of it, appear to perform better then PCA under most circumstances.

Other subspace methods have also been exploited for the face recognition task: Foon et al. [184] have integrated various wavelet transforms and non-negative matrix factorizations [185] and claim to have obtained better verification rates as compared to the basic eigenfaces approach. In [186], an intra-class subspace is constructed, and the classification is based on the nearest weighted distance between the query face and each intra-class subspace. Experimental results are presented to demonstrate that this method performs better than some other nearest feature techniques.

A study and comparison of four subspace representations for face recognition, i.e., PCA, ICA, Fisher Discriminant Analysis (FDA) and probabilistic eigenfaces and their ‘kernalized’ versions (if available), is presented in [187]. A comprehensive review of recent advances in subspace analysis for face recognition can be found in [188].

2.5.1.2.2. AI. AI approaches utilize tools such as neural networks and machine learning techniques to recognize faces. Some examples of methods belonging to this category are given below.

In [189], 50 principal components were extracted and an autoassociative neural network was used to reduce those components to five dimensions. A standard multi-layer perceptron was exploited to classify the resulting representation. Though favorable results were received, but the database used for training and testing was quite simple: the pictures were manually aligned, there was no lighting variation, tilting, or rotation, and there were only 20 people in the database.

Weng et al. [190] made use of an hierarchical neural network which was grown automatically and not trained on the traditional gradient descent method. They reported good results on a database of 10 subjects.

Lawrence et al [10] reported a 96.2% recognition rate on the ORL database (a database of 400 images of 40 individuals) using a hybrid neural network solution which combines local image sampling, a self-organizing map [191, 192] neural network (which provides dimensionality reduction and invariance to small changes in the image sample) and a convolutional neural network (which provides partial invariance to translation, rotation, scale and deformation). The eigenfaces method [6, 123] produced a 89.5% recognition accuracy on the same data. Replacing the self-organizing map by Karhunen-Loeve transform and the convolutional network by a multi-layer perceptron resulted in a recognition rate of 94.7% and 60% respectively (Figure 2.13).

Eleyan and Demirel [193] used principal components analysis to obtain feature projection vectors from face images which were then classified using feed forward neural networks. Some tests on the ORL database using various number of training and testing images showed that the

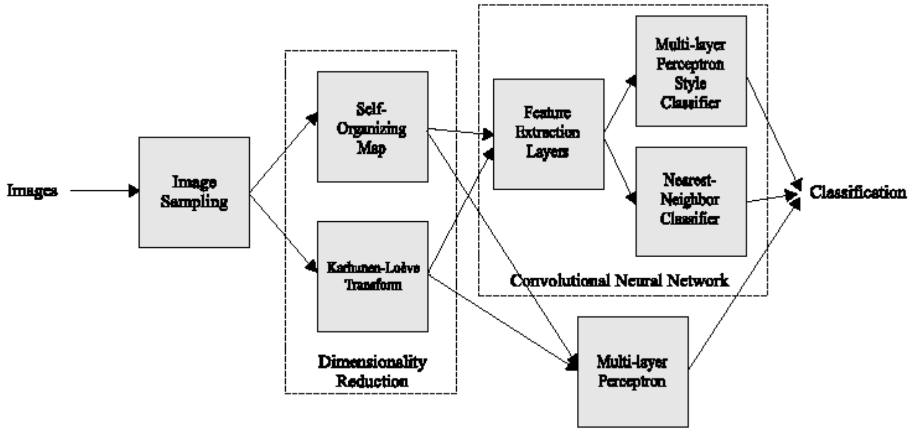


Figure 2.13. A high-level diagram of the system used for face recognition [10]. (©1997 IEEE)

performance of this system was better than the eigenfaces [6, 123] one in which a nearest neighbor classifier was used for classification.

Li and Yin [194] introduced a system in which a face image is first decomposed with a wavelet transform to three levels. The Fisherfaces method [8] is then applied to each of the three low-frequency sub-images. Then, the individual classifiers are fused using the RBF neural network. The resulting system was tested on images of 40 subjects from the FERET database and was shown to outperform the individual classifiers and the direct Fisherfaces method.

Melin et al. [195] divided the face into three regions (the eyes, the mouth and the nose) and assigned each region to a module of the neural network. A fuzzy Sugeno integral was then used to combine the outputs of the three modules to make the final face recognition decision. They tested it on a small database of 20 people and reported that the modular network yielded better results than a monolithic one.

Recently, Zhang et al. [196] have proposed an approach in which a similarity function is learned describing the level of confidence that two images belong to the same person, similar to [127]. The facial features are selected by obtaining Local Binary Pattern (LBP) [197] histograms of sub-regions of the face image and the Chi-square distances between corresponding LBP histogram are chosen as the discriminative features. The AdaBoost learning algorithm, introduced by Freund and Schapire [198], is then applied to select the most efficient LBP features as well as to obtain the similarity function in the form of linear combination of LBP feature based weak learners. Experimental results on the FERET frontal image sets has shown that this method yields a better recognition rate of 97.9 % by utilizing less features than a previous similar approach proposed by Ahonen et al. [199].

Some researchers have also used the one-against-one approach [200] for decomposing the multi-class face recognition problem into a number of binary classification problems. In this method, one classifier is trained for each pair of classes, ignoring all the remaining ones. The outputs of all the binary classifiers are then combined to construct the global result. For binary classifiers with probabilistic outputs, *pairwise coupling* (PWC) [201] can be used to couple these outputs into a set of posterior probabilities. Then, the test example is assigned to the class with the maximum posterior probability. One main disadvantage of PWC is that when a test example does not belong to either class related to a binary classifier, then the output of that classifier is meaningless and can damage the global result. In [202], a new algorithm called PWC-CC (where CC stands for correcting classifier) is presented to solve this problem: for each binary classifier separating class c_i from class c_j, a new classifier separating the two classes from all other classes is trained. Even though PWC-CC performs better than PWC, but it has its own drawbacks. In [203], a novel PWC-CC (NPWC-CC) method has been suggested for the face recognition problem and results of tests on the ORL database have been presented to support the claim that it outperforms PWC-CC. In [204], the optimal PWC (O-PWC) approach has been introduced and it is shown to have better recognition rates that the PWC method. Feature extraction is done by using principal components analysis in [203] and by wavelet transform in [204]. In both [203] and [204], Support Vector Machines (SVMs) have been employed as binary classifiers and the SVM outputs have been mapped to probabilities by using the method suggested by Platt [205].

It should be noted that Support Vector Machine (SVM) is considered to be one of the most effective algorithms for pattern classification problems [206]. In general, it works as follows for binary problems [207]: First, the training examples are mapped to a high dimensional feature space *H*. Then, the optimal hyperplane in *H* is sought to separate examples of different classes as much as possible, while maximizing the distance from either class to the hyperplane. SVM has been employed for face recognition by several other researchers and has been shown to yield good results [206, 208-212].

Hidden Markov models [213] have also been employed for the face recognition task. Samaria and Harter [214] used a one-dimensional HMM to obtain a peak recognition accuracy of 87% on the ORL database. They later upgraded the one-dimensional HMM to a pseudo two-dimensional HMM [215] and achieved a best recognition performance of 95% on the same database using half the images for training and the other half for testing. Nefian and Hayes III [216] reported a best recognition rate of 98% on the same training and testing sets using embedded HMM [217] face models and they also claimed that their system was much faster than that of Samaria [215] and was invariant to the scale of the face images.

Some other AI approaches utilized for the face recognition task include evolutionary pursuit [218, 219] and techniques [220, 221] based on boosting [198, 222]. These schemes have reportedly yielded promising results for various difficult face recognition scenarios.

2.5.1.2.3. Multiple classifier systems. Since the performance of any classifier is more sensitive to some factors and relatively invariant to others, a recent trend has been to combine individual classifiers in order to integrate their complementary information and therefore, create a system that is more robust than any individual classifier to variables that complicate the recognition task. Such systems have been termed as multiple classifier systems (MCSs) [223] and are a very active research area at present. Examples of such approaches proposed for face recognition include the following: Lu et al. [224] fused the results of PCA, ICA and LDA using the sum rule and RBF network [225] based integration strategy (Figure 2.14); Marcialis and Roli [12, 226, 227] combined the results of the PCA and LDA algorithms (Figure 2.15); Achermann and Bunke [228] utilized simple fusion rules (majority voting, rank sum, Baye’s combination rule) to integrate the weighted outcomes of three classifiers based on frontal and profile views of faces; Tolba and Abu-Rezq [229] employed a simple combination rule for fusing the decisions of RBF and LVQ networks; Wan et al. [230] used a SVM and HMM hybrid model; Kwak and Pedrycz [231] divided the face into three regions, applied the Fisherfaces method to the regions as well as the whole face and then integrated the classification results using the Choquet fuzzy integral [232]; Haddadnia et. al. [233] used PCA, Pseudo Zernike Moment Invariant (PZMI) [234, 235] and Zernike Moment Invariant (ZMI) for extracting feature vectors in parallel which were then classified simultaneously by separate RBF neural networks and the outputs of these networks were then combined by a majority rule to determine the final identity of the individual in the input image.

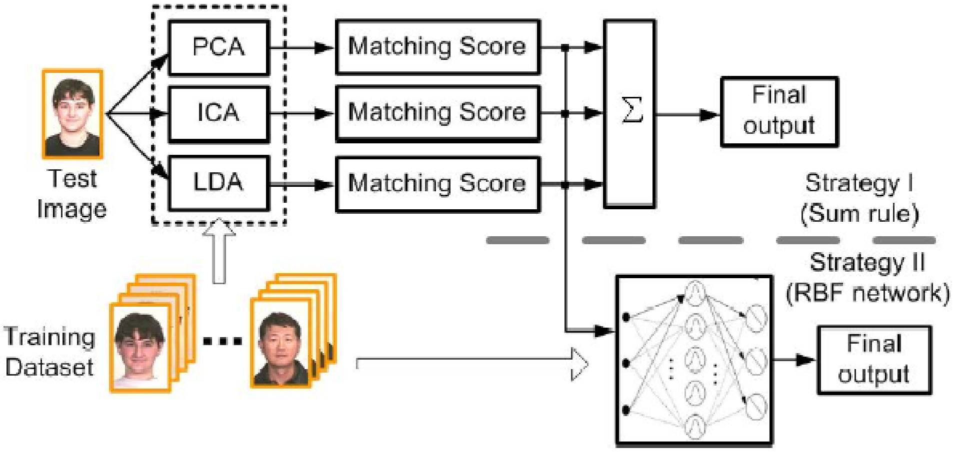


Figure 2.14. Classifier combination system framework [224]. (©2003 IEEE)

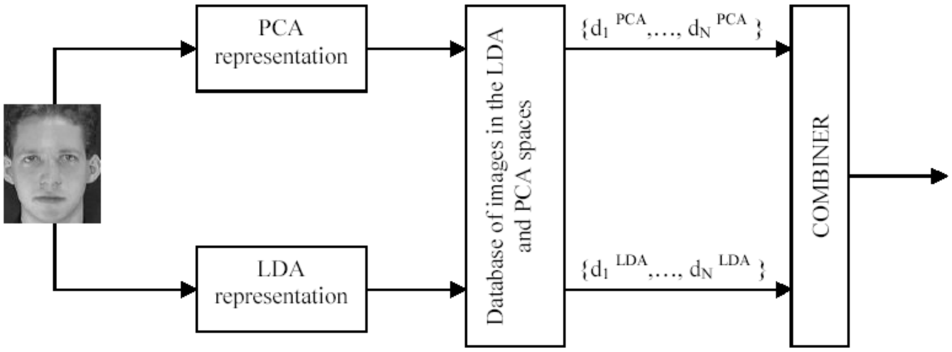


Figure 2.15. Overview of the fusion methodology [12]. (With kind permission of Springer Science and Business Media)

2.5.1.2.4 Advantages and disadvantages. The main advantage of the holistic approaches is that they do not destroy any of the information in the images by concentrating on only limited regions or points of interest [83]. However, as mentioned above, this same property is their greatest drawback, too, since most of these approaches start out with the basic assumption that all pixels in the image are equally important [117]. Consequently, these techniques are not only computationally expensive but require a high degree of correlation between the test and training images and do not perform effectively under large variations in pose, scale and illumination etc. [17]. Nevertheless, as mentioned in the above review, several of these algorithms have been modified/enhanced to compensate for such variations and dimensionality reduction techniques have been exploited (note that even though such techniques increase generalization capabilities, the down side is that they may potentially cause the loss of discriminative information [188]), as a result of which these approaches appear to produce better recognition results than the feature-based ones in general. In the latest comprehensive FERET evaluation [105, 106], the probabilistic eigenface [127], the Fisherface [8] and the EBGM [102] methods were ranked as the best three techniques for face recognition (Even though the EBGM method is feature-based in general, but its success depends on its application of holistic neural network methods at the feature level).

2.5.2. Face recognition from video sequences

Since one of the major applications of face recognition is surveillance for security purposes which involves real-time recognition of faces from an image sequence captured by a video camera, therefore, a significant amount of research has been directed towards this area in recent years.

A video-based face recognition system typically consists of three modules: one for detecting the face, a second one for tracking it and a third one for recognizing it [236]. Most of these systems choose a few good frames and then apply one of the recognition techniques for intensity images to those frames in order to identify the individual [237]. A few of these approaches are briefly described below.

Howell and Buxton [238] employed a two-layer RBF network [239, 240] for learning/training and used Difference of Gaussian (DoG) filtering and Gabor wavelet analysis for the feature representation while the scheme from [241] was utilized for face detection and tracking. Training and testing was done using two types of image sequences: 8 primary sequences taken in a relatively constrained environment and a secondary sequence recorded in a much more unconstrained atmosphere (Figures 2.16, 2.17). The image sequences consisted of 62 to 94 frames. The use of Gabor wavelet analysis for feature representation, as opposed to DoG filtering, seemed to yield better recognition results. The recognition accuracies reported varied

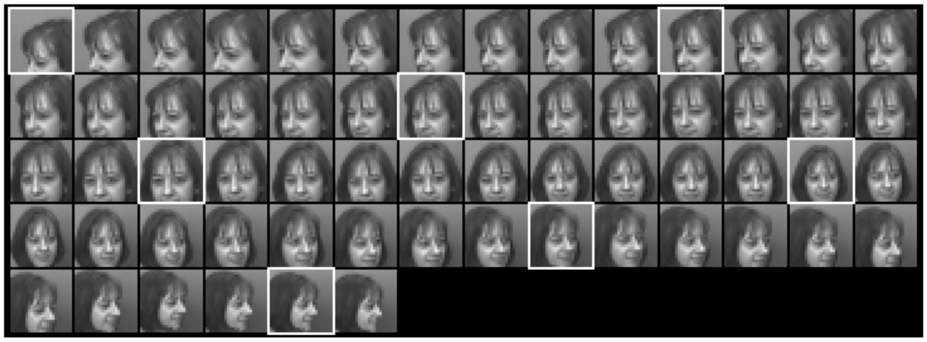


Figure 2.16. A complete Primary sequence for the class *Carla*, after segmentation but before preprocessing [238]. (©1996 IEEE)

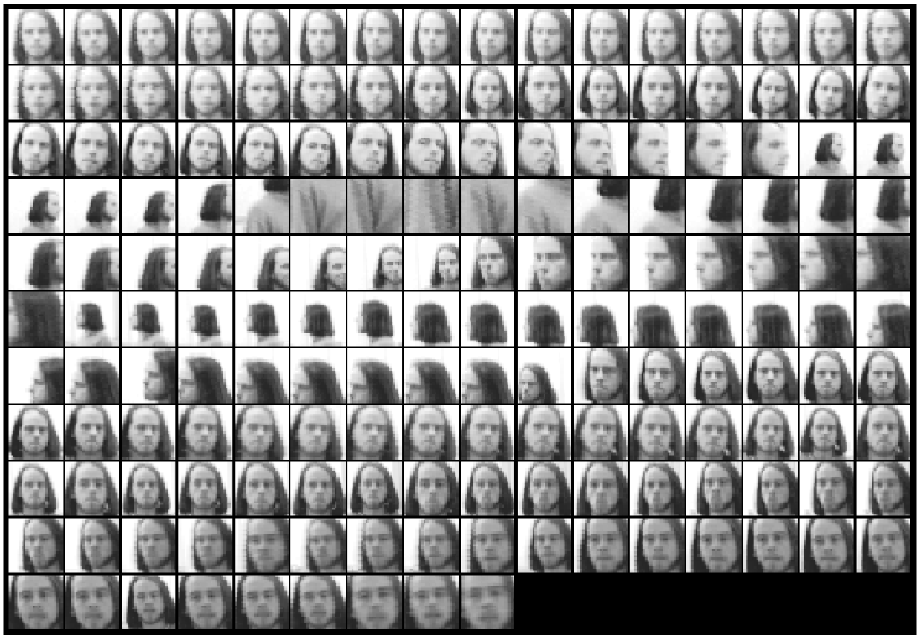


Figure 2.17. A complete Secondary sequence for the class *steve*, after segmentation but before preprocessing [238]. (© 1996IEEE)

quite a bit, ranging from 99%, using 278 images for training and 276 for testing, to 67%, using 16 training and 538 testing images.

de Campos et al. [242] propose a recognition system which uses skin color modeling [243] to detect the face; then utilizes GWN [244] to detect prominent facial landmarks (i.e., the eyes, nose and mouth) and to track those features. For each individual frame, eigenfeatures [1] are then extracted and a feature selection algorithm [245] is applied over the combination of all the eigenfeatures and the best ones are selected to form the feature space. A couple of classifiers described in [246] are then applied to identify the individual in the frame and finally, a super-classifier based on a voting scheme [247] performs the final classification for the entire video sequence (Figures 2.18, 2.19). Good recognition results (97.7% accuracy) have been reported using 174 images of eyes of 29 people (6 images per person).

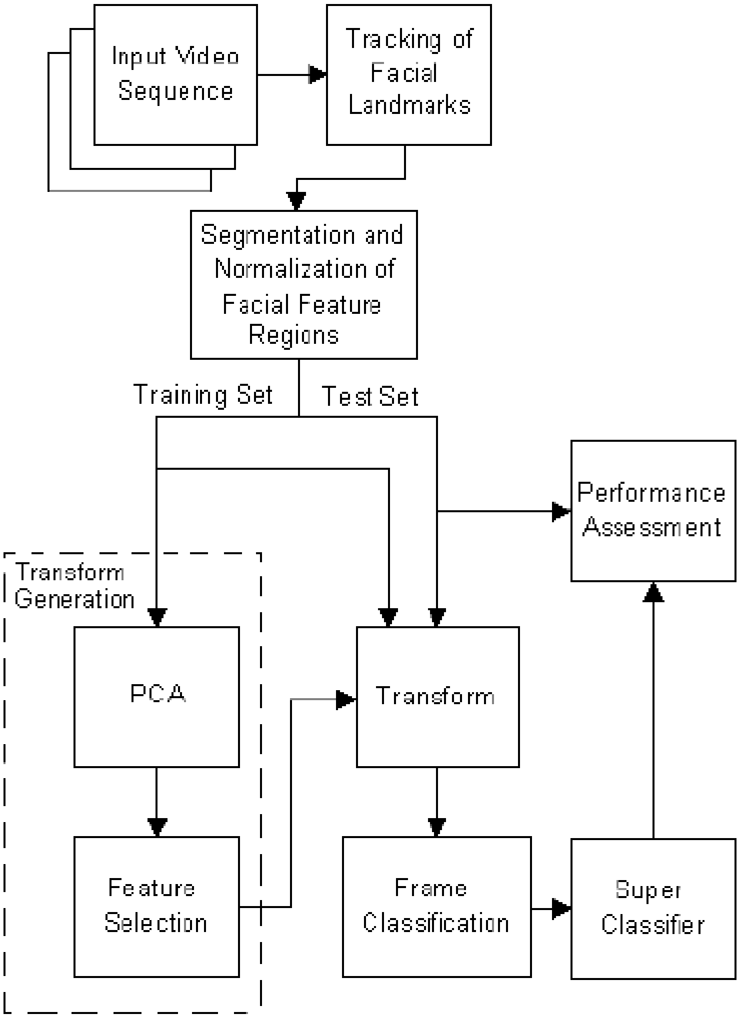


Figure 2.18. Overview of the project [242]. (Courtesy of T. E. de Campos, R. S. Feris and R. M. Cesar Jr.)

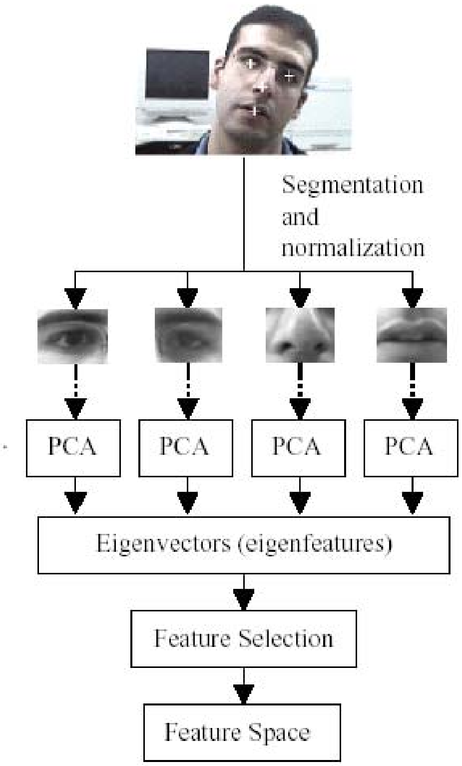


Figure 2.19. Feature space generation [242] (Courtesy of T. E. de Campos, R. S. Feris and R. M. Cesar Jr.)

Biuk and Loncaric [248] take image sequences in which the pose of the person's face changes from -90 degrees to 90 degrees (Figure 2.20). The sequences are projected into eigenspace to give a prototype trajectory for each known individual. During the recognition phase, an unknown face trajectory is compared with the prototype trajectories to determine the identity of the individual (Figure 2.21). They tested the system on a database of 28 individuals (11 frames per person) and found that the system yielded excellent recognition results when all frames and 4 or more eigenfaces were used but the performance decreased when either parameter was decreased. They, however, propose a matching method which associates a score to each trajectory point and then makes the final match based on the maximum score and they claim that this enhancement enables their system to achieve better performance when a smaller number (< 4) of eigenfaces are used.

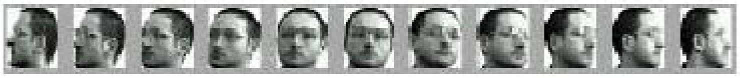


Figure 2.20. Face images taken from different view angles (profile to profile) [248]. (©2001 IEEE)

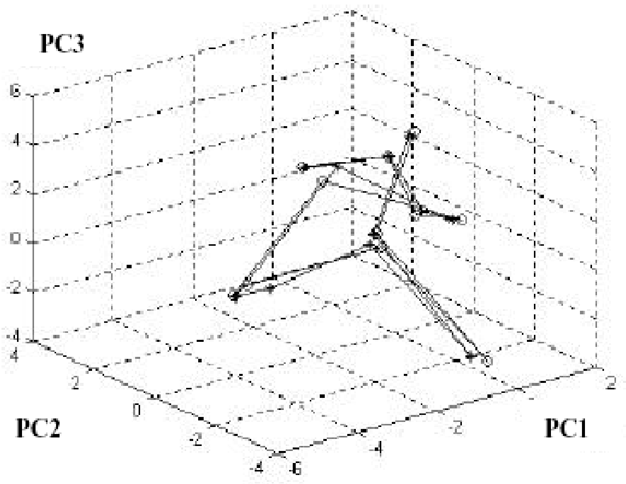


Figure 2.21. Pattern trajectories for two sequences of the same person represented with first three principal components [248]. (©2001 IEEE)

Recently, some approaches [249, 250] have employed a video-to-video paradigm in which information from a sequence of frames from a video segment is combined and associated with one individual. This notion involves a temporal analysis of the video sequence and a condensation of the tracking and recognition problems. Such schemes are still a matter of ongoing research since the reported experiments were performed without any real variations in orientation and facial expressions [251].

It is worth mentioning that several schemes have incorporated information from other modalities in order to recognize facial images acquired from video clips. For instance, [252] makes use of stereo information and reports a recognition accuracy of 90% while [55] exploits both audio and video cues as well as 3D information about the head to achieve 100% accuracy rate for 26 subjects. (For more information about recognition based on audio and video, see the Proceedings of the AVBPA Conferences [253]).

A detailed survey of recent schemes for face recognition from video sequences is provided in [254].

2.5.2.1. Advantages and disadvantages. Dynamic face recognition schemes appear to be at a disadvantage relative to their static counterparts in general since they are usually hampered by one or more of the following: low quality images (though image quality may be enhanced exploiting super-resolution techniques [255-258]), cluttered backgrounds (which complicate face detection [259]), more than one face in the picture and a large amount of data to process [114]. Furthermore, the face image may be much smaller than the size required by most systems employed by the recognition modules [237].

However, dynamic schemes do have the following advantages over static techniques: the enormous abundance of data empowers the system to choose the frame with the best possible image and discard less satisfactory ones [238]. Video provides temporal continuity [238], so classification information from several frames can be combined to improve recognition performance. Moreover, video allows tracking of face images such that variations in facial expressions and poses can be compensated for, resulting in improved recognition [38]. Dynamic schemes also have an edge over static ones when it comes to detecting the face in a scene since these schemes can use motion to segment a moving person's face [114].

2.5.3. Face recognition from other sensory inputs

Though the bulk of the research on face recognition has been focused on identifying individuals from 2D intensity images but nevertheless, in recent years, some attention has also been directed towards exploiting other sensing modalities, such as 3D or range data and infra-red imagery, for this purpose.

2.5.3.1. 3D model-based. The main argument in favor of using 3D information for face recognition appears to be that it allows us to exploit features based on the shape and the curvature of the face (such as the shape of the forehead, jaw line, and cheeks) without being plagued by variances caused by lighting, orientation and background clutter that affect 2D systems [83, 260, 261]. Another argument for the use of depth data is that "at our current state of technology it is the most straightforward way to input or record complex shape information for machine analysis" [262]. The obvious drawbacks of such approaches are their complexity and computational cost [237].

The following techniques are currently being used to obtain 3D information [260]:

- Scanning systems: Laser face scanners produced by companies like Cyberware Inc. [263] and 3D Scanners Ltd. [264] seem to be producing highly accurate results; however, the cost of these commercial scanning services is obviously substantial.

- Structured light systems: These systems make use of the principles of stereo vision to obtain the range data. Their main advantage is that the only equipment they require is cameras and some kind of projection system. The primary drawback of such systems is that they can have difficulty with resolving the shape of the pattern in the camera image.
- Stereo vision systems: These are systems that attempt to extract 3D information from two or more 2D images of the same object taken from different angles. They are limited to objects which will "generate a sufficient number of image features to allow for conclusive stereo matching. In the case of trying to establish the shape of a reasonably smooth object, such as the human face, these systems would be unable to generate accurate surface shape. (Smooth surfaces can be 'roughed up' by the projection of a textured pattern onto the face)" [260].
- Reverse rendering/shape from shading: These techniques endeavor to construct the shape of an object using knowledge about illumination and the physical properties of the object.

Until recently, there appear to have been very few papers that describe attempts to recognize faces based mainly on range data or 3D data about the subjects' faces. However, lately, there has been revival of interest in this area and several new schemes of this sort have been proposed in the past few years. One of the earliest of such approaches is described in [262] where the principle curvatures of the face surface are calculated from range data (Figure 2.22) and then this data, supplemented by a priori information about the structure of the face, is used to locate the various facial features (i.e., the nose eyes, forehead, neck, chin etc.). The faces are then normalized to a standard position and reinterpolated onto a regular cylindrical grid. The volume of space between two normalized surfaces is used as a similarity measure. The system was tested using face images of 8 people (3 images per person). Features were detected adequately for all

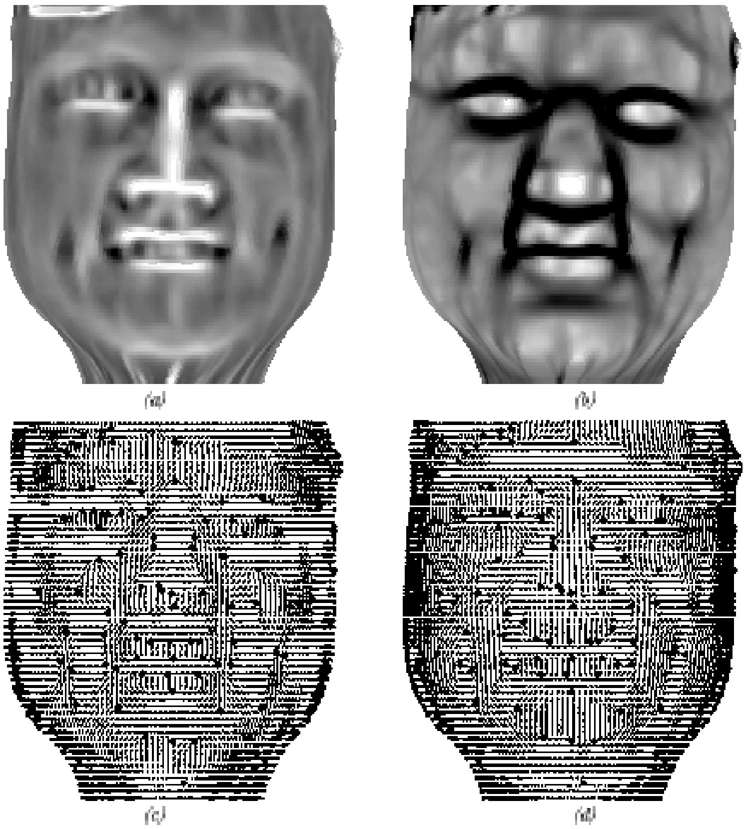


Figure 2.22. Principle curvatures for a single face: magnitude (a) and direction (c) of maximum curvature, magnitude (b) and direction (d) of minimum curvature. Umbilic points are marked in (c) and (d); filled circles are points with positive index and open circles are points with negative index [262]. (Courtesy of G. Gordon)

faces. Recognition rates of 97% and 100% were reported for individual features and the whole face respectively.

Another approach described in [17] uses profiles (which have external contours consisting of rather rigid parts) instead of frontal images, captures 3D data by triangulation and then does a 3D comparison of the profile data (Figures 2.23, 2.24). This method requires a good deal of user cooperation and restrictions on the background etc. in order for it to work.

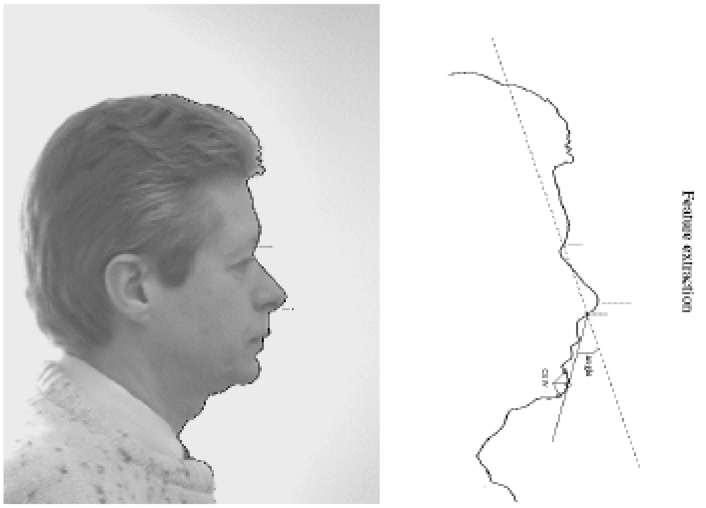


Figure 2.23. (a) Profile image (b) Contour extraction, nose and eye localization, feature extraction [17]. (Courtesy of C. Beumier and M. Acheroy)

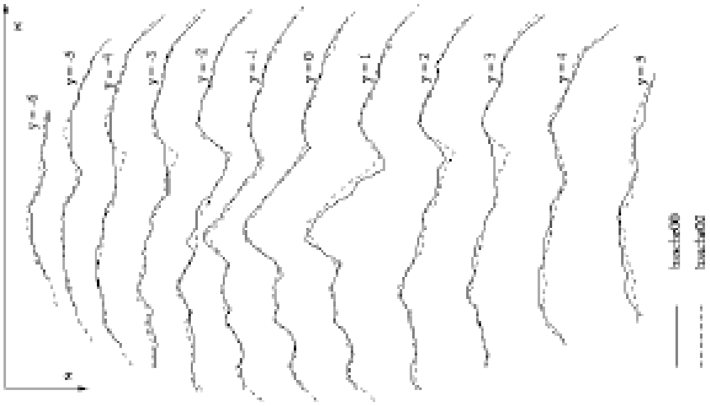


Figure 2.24. 3D comparison by parallel planar cuts [17]. (Courtesy of C. Beumier and M. Acheroy)

[83] uses 3D data to normalize the results from the face detection algorithm to a form more appropriate for the recognition engine, i.e., in this case, the 3D data is just being used to supplement rather than supplant existing face detection and recognition algorithms.

Some examples of more recent face recognition approaches based on 3D data include the following: Castellani et al. [265] approximate range images of faces obtained by stereoscopic analysis using Multi-level B-Splines [266] and SVMs are then used to classify the resulting approximation coefficients. Some other techniques [261, 267, 268] first project the 3D face data into a 2D intensity image and the projected 2D images are then processed as standard intensity images. Yet other methods have been proposed for 3D face recognition based on local features [269], local and global geometric cues [270], profiles [271-274] and rank-based decision fusion of various shape-based classifiers [275].

Several approaches have also been proposed that integrate 2D texture and 3D shape information. Such methods make use of PCA of intensity images [276-278], facial profile intensities [279], Iterative Closest Point (ICP [280]) [281, 282], Gabor wavelets [283] and Local Feature Analysis [284], etc. For instance, Wang et al. [283] extract 3D shape templates from range images and texture templates from grayscale images of faces, apply PCA separately to both kinds of templates to reduce them to lower-dimensional vectors, then concatenate the shape and texture vectors and finally apply SVMs to the resulting vectors for classification. In general, experiments with such systems indicate that combining shape and texture information reduces the misclassification rates of the face recognizer.

Comprehensive recent surveys of literature on 3D face recognition can be found in [285] and [286].

2.5.3.2. Infra-red. Since thermal infra-red imagery of faces is relatively insensitive to variations in lighting [287], hence, such images can be used as an option for detecting and recognizing faces. Furthermore, [288] argues that since infra-red facial images reveal the vein and tissue structure of the face which is unique to each individual (like a fingerprint), therefore, some of the face recognition techniques for the visible spectrum should yield favorable results when applied to these images. However, there exist a multitude of factors that discourage the exploitation of such images for the face recognition task, among them, the substantial cost of thermal sensors, the low resolution and high level of noise in the images, the lack of widely available data sets of infra-red images, infra-red radiation's being opaque to glass (making it possible to occlude part of the face by wearing eyeglasses) [289] and last, but not least, the fact that infra-red images are sensitive to changes in ambient temperature, wind and metabolic processes in the subject [290] (Note that in [291], the use of blood perfusion data is suggested to alleviate the effect of ambient temperature).

In [288], the eigenface technique [6, 123] was applied to a database of 288 hand-aligned low-resolution(160x120) images of 24 subjects taken from 3 viewpoints. The following recognition rates were reported: 96% for frontal views, 96% for 45 degrees views, and 100% for profile views.

Wilder et al. [292] compared the performance of three face recognition algorithms on a database of visible and infra-red images of 101 subjects and concluded that the recognition results for one modality were not significantly better than those for the other.

Socolinsky et al. [18] tested the eigenfaces [6, 123] and the ARENA [293] algorithms on a database of visible and infrared images of 91 distinct subjects (captured under various illumination conditions, facial expressions and with or without glasses using a sensor capable of

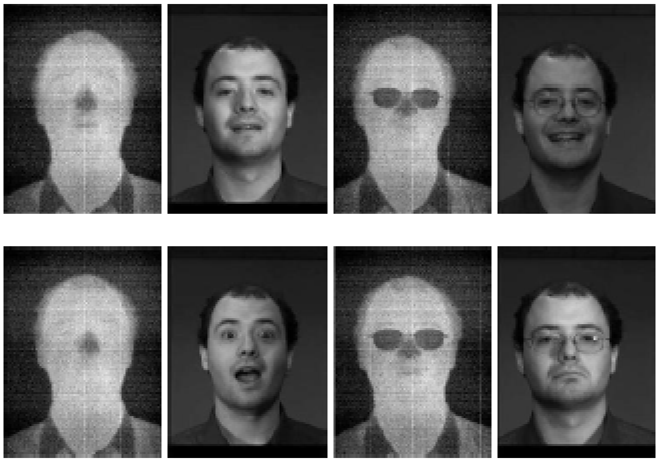


Figure 2.25. Sample imagery from the database. Note that LWIR images are not radiometrically calibrated [18]. (©2001 IEEE)

Figure 2.26. First five visible eigenfaces [18]. (© 2001[18] IEEE)

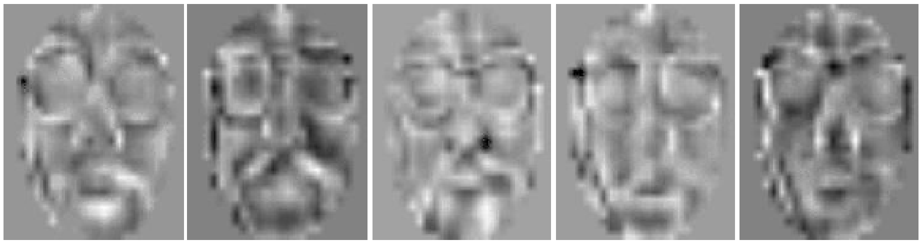


Figure 2.27. First five LWIR eigenfaces [18]. (©2001 IEEE)

imaging both modalities simultaneously (Figures 2.25, 2.26 and 2.27)) and reported that the infra-red imagery significantly outperformed the visible one in all the classification experiments under the various above-mentioned conditions.

Selinger and Socolinsky [290] used the same database of 91 subjects and tested the performance of four face recognition algorithms (PCA, LDA, LFA and ICA) under the aforementioned conditions and reached the same conclusion though they did concede that the apparent superiority of the infra-red approach may stem from the fact that their data did not contain sufficiently challenging situations (i.e., changes in temperature, wind etc.) for the infra-red imagery whereas it did so for the visible images.

Chen et al. [294] collected several datasets of images (both in infra-red and the visible spectrum) of 240 distinct subjects under various expressions and lighting conditions at different times (some of the images were taken in the same session while others were taken over a period of ten weeks (Figure 2.28)). They studied the effect of temporal changes in facial appearance on the performance of the eigenfaces algorithm in both modalities and concluded that though the recognition accuracy was approximately the same for images taken in the same session, visible images outperformed the infra-red ones when there was a significant lapse in the time between which the training and test images were acquired. They attributed the lower performance of the infra-red imagery to variations in the thermal patterns of the same subject and the sensitivity of the infra-red imagery to the manual location of the eyes. They also found that the FACEIT [295] software performed better than eigenfaces in both modalities. However, the combination of the two classifiers using the sum rule [296] outperformed the individual classifiers as well as the FACEIT program. Latter experiments conducted by Chen et al. [20] on a larger set of data reconfirmed these results.

(a) Week 1 (b) Week 2

(a) Week 3 (b) Week 4

(a) Week 5 (b) Week 6

(a) Week 7 (b) Week 8

(a) Week 9 (b) Week 10

Figure 2.28. Normalized face images of one subject in visible and IR across ten weeks [20]. (Reprinted from Computer Vision and Image Understanding, 99(3), X. Chen, P. Flynn and K. Bowyer, IR and Visible Light Face Recognition, 332-358, Copyright (2005), with permission from Elsevier)

Other approaches have also been proposed recently which explore the fusion of face images captured under visible and infrared light spectrum to improve the performance of face recognition [289, 297-299].

A comprehensive review of recent advances in face recognition from infra-red imagery may be found in [300].

2.6. Conclusions

Face recognition is a challenging problem in the field of image analysis and computer vision that has received a great deal of attention in the past several years because of its many applications in various domains. Research has been conducted vigorously in this area for the past four decades or so and though huge progress has been made, encouraging results have been obtained and current face recognition systems have reached a certain degree of maturity when operating under constrained conditions, however, they are far from achieving the ideal of being able to perform adequately in all the various situations that are commonly encountered by applications utilizing these techniques in practical life. The ultimate goal of researchers in this area is to enable computers to emulate the human vision system and, as has been aptly pointed out by Torres [38], "Strong and coordinated effort between the computer vision, signal processing and psychophysics and neurosciences communities is needed" to attain this objective.

CHAPTER 3

ANALYSIS OF SUBSPACE-BASED FACE RECOGNITION TECHNIQUES UNDER CHANGES IN IMAGING FACTORS[1]

---

[1] Rabia Jafri and Hamid R. Arabnia. 2007. Proceedings of the 4th International Conference on Information Technology - New Generations (ITNG 2007; Data Mining Track), April 2-4, Las Vegas, USA; IEEE Computer Society. 406-413.
Reprinted here with permission of publisher.

Abstract

The recognition accuracy of three face recognition algorithms based on subspace methods (i.e., the Eigenfaces approach, the Fisherfaces technique, and the probabilistic Eigenfaces approach) is tested for different gallery sizes, image resolutions and number of bits per pixel. The same experiments are then repeated for various combinations of these techniques and the performance of these combinations with that of the individual methods is compared. The results reveal some interesting relations between recognition accuracy and image resolution and also indicate that the combinations of these techniques do yield some improvement in performance over the individual methods warranting that it would be worthwhile to further investigate the unification of these approaches.

3.1. Introduction

Face recognition is a challenging problem in the field of computer vision which has received a good deal of attention lately because of its numerous applications in domains such as surveillance, security and adaptive human-computer interfaces. Though several methods have been proposed for this task, the techniques based on subspace methods, in particular, have become the focus of extensive research in recent years [301]. These approaches apply statistical dimensionality reduction techniques to a small set of representative face images to extract the most meaningful dimensions, which characterize the subspace in which faces are located, from the entire image space. The set of images to be recognized are then projected onto this new subspace and classified using a nearest neighbor classifier. The popularity of these schemes stems from the advantages they offer in terms of their relatively fast speed, compact representation, ease of implementation, and non-stringent data requirements (i.e., they do not require full 3D models or detailed geometry). Furthermore, they appear to perform well in general in relatively constrained environments. Two of the most popular of such schemes, the probabilistic Eigenfaces method [128, 302], and the Fisherfaces method [303, 304], were ranked among the best techniques for face recognition in the most recent FERET evaluation [305].

Though much work has been done in the past few years to test the robustness of subspace-based face recognition methods to changes in factors like illumination, pose and scale [306], however, factors like imaging parameters (i.e., image resolution and number of bits per pixel), have not been studied in as much detail. We believe that an extensive study into the effects of these factors on recognition accuracy is warranted for the following reasons, among others: 1) Applications with real-time speed requirements benefit from using low resolution images with fewer bits per pixel. The said study would help surmise what could be the lowest resolution images with the fewest bits per pixel that can be used without compromising the recognition accuracy. 2) Such a study may reveal that methods which perform well using high resolution images with more bits per pixel suffer a substantial drop in performance at lower resolutions and fewer bits per pixel; the reverse might be true for some other techniques. Thus, if one is forced to use images of a certain resolution and number of bits per pixel, the study would help choose the method that performs the best for that scenario.

A study to test the effects of varying the resolution and number of bits per pixel of the face images on three well-know face recognition algorithms has, therefore, been undertaken in this paper. The algorithms chosen are the probabilistic Eigenfaces method [128, 302], the Fisherfaces method [303, 304] and the basic Eigenfaces method [6]. Moreover, motivated by the recent interest in Multiple Classifier Systems [223] (i.e., systems that integrate the complementary information of individual classifiers to create a system that is more robust to variables that complicate the recognition task), the results obtained by fusing the outcomes of these algorithms in various ways have also been examined.

The rest of this paper is organized as follows: Section 3.2 contains a brief overview of the face recognition algorithms mentioned above. Section 3.3 describes the tests conducted to evaluate and compare the performance of these algorithms and various combinations thereof as well as an analysis of the outcomes of these experiments. Section 3.4 contains some concluding remarks.

3.2. Face recognition algorithms utilizing subspace methods

A brief overview of the face recognition algorithms chosen for experimentation is presented below.

3.2.1. Eigenfaces

A face image represented as a t-dimensional vector can be thought of as a point in a t-dimensional space. Since face images are so similar to each other in their overall configuration, they will form a tight cluster in this t-dimensional image space. The basic idea behind Principal Components Analysis (PCA), when applied to the face recognition problem, is to find the vectors in this vast image space which best account for the variation among a given collection of face images. These vectors define the subspace in which the face images are clustered. These vectors are simply the eigenvectors of the eigensystem of the covariance matrix of the given set of face images and can be found by singular value decomposition (See [6] for details).

The eigenvectors are then sorted in descending order of their associated eigenvalues and a certain number of those with the lowest eigenvalues are dropped (since it is assumed that these vectors tend to encode noise). The images of known individuals are then projected onto the retained eigenvectors to form face class vectors. The input image to be recognized is also projected onto the retained eigenvectors and the resulting projection vector is then compared to each of the face class vectors using a similarity criteria or distance measure of choice to find the best match.

3.2.2. Fisherfaces

Fisher's Linear Discriminant (FLD) [125] is a class-specific method that aims to produce a linear transformation that stresses the differences between classes while minimizing the differences within them. The goal is to create a subspace that is linearly separable between classes. Note that in the context of the face recognition problem, a class refers to an individual and the samples in a class are the face images of that individual.

Let S_B be the between-class scatter matrix defined as

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad \textbf{(3.1)}$$

and S_W be the within-class scatter matrix defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \qquad \textbf{(3.2)}$$

where μ =mean image of all samples, μ_i=mean image of class X_i, N_i= number of samples in class X_i, c=number of classes. Then a matrix W which maximizes the function $J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}$, can be found by solving the generalized $S_B w_i = \lambda_i S_w w_i$ eigensystem [307]. More specifically, W consists of the eigenvectors associated with the k largest eigenvalues of the above system. To ensure that S_W is non-singular, usually PCA is applied to reduce the dimensionality of the training images before applying the FLD procedure. The PCA and FLD basis vectors are then multiplied to get the final transformation matrix *W*.

The recognition process for Fisherfaces is the same as for the Eigenfaces algorithm except that the eigenvectors now being used are those obtained by the training procedure described above (i.e., the ones that *W* is comprised of) [308].

3.2.3. Bayesian Intrapersonal/Extrapersonal Classifier (probabilistic Eigenfaces approach)

Moghaddam and Pentland [302] propose an alternative approach which utilizes difference images, where a difference image for two face images is defined as the signed arithmetic difference between the intensity values of the corresponding pixels in those images. Two classes of difference images are defined: *intrapersonal*, which consists of difference images originating from two images of the same person, and *extrapersonal,* which consists of difference images derived from two images of different people.

It is assumed that both these classes originate from discrete Gaussian distributions within the space of all possible difference images. The actual parameters of these distributions are not known and are estimated from the training data by applying PCA to the intrapersonal and extrapersonal subspaces. Then, given the difference image between two images I_1 and I_2, the probability that the difference image belongs to the intrapersonal class is given by Bayes Rule as follows:

$$P(\Omega_I \mid d(I_1, I_2)) = \frac{P(d(I_1, I_2) \mid \Omega_I)P(\Omega_I)}{P(d(I_1, I_2) \mid \Omega_I)P(\Omega_I) + P(d(I_1, I_2) \mid \Omega_E)P(\Omega_E)} \quad \textbf{(3.3)}$$

where $d(I_1, I_2)$ = the difference image between two images I_1 and I_2, $\mathbf{\Omega}_I$ = the intrapersonal class, $\mathbf{\Omega}_E$= the extrapersonal class.

Classification is performed using the maximum a posteriori (MAP) rule – the two images are declared to belong to the same individual if $P(\Omega_I|d(I_1,I_2)) > P(\Omega_E|d(I_1,I_2))$ or equivalently, if $P(\Omega_I|d(I_1, I_2)) > ½$. For a computationally more expedient approach, Moghaddam and Pentland [128] also suggest ignoring the extrapersonal class information and calculating the similarity based only on the intrapersonal class information. In the resulting maximum likelihood (ML) classifier, the similarity score is given only by $P(d(I_1,I_2)|\Omega_I)$.

3.3. Evaluation and comparison of subspace-based face recognition algorithms

3.3.1. Related work

Much research has been conducted in the past several years to examine and compare the performance of the afore-mentioned algorithms as well as other subspace-based algorithms. The landmark in such testing was the FERET evaluation [305] which tested various algorithms on large databases of images and different poses and also evaluated how well they rejected faces not in the database of known individuals. Subsequently, several independent studies have been carried out to evaluate such techniques, a few examples being [301, 304, 308]. Specific aspects of these algorithms, such as the use of various distance measures as similarity metrics [309], as well as the application of these techniques to different sensory inputs, such as infra-red imagery [300], have also been studied. However, to the best of our knowledge, there appears to be a lack of any research to investigate how the performance of these algorithms, or any combination thereof, varies if the resolution of the face images or the number of bits per pixel is modified. The only work in this regard appears to be done by Wang et al [48], who report that recognition accuracy for the Eigenfaces and Fisherfaces algorithm increases with increasing image resolution but becomes stable at resolution 64x48. However, they do not vary the gallery size and all tests are conducted on a fixed gallery of 126 subjects. Experimentation to examine these hitherto unexplored factors is, therefore, undertaken in this paper.

3.3.2. Experimentation

The following terminology introduced in [305] will be used in the rest of the paper: A *training set* is a collection of images used to construct the subspace; a *gallery* is a set of images of known individuals; a *probe set* consists of face images that need to be identified.

3.3.2.1. Obtaining the data (face images) and code. The data for these experiments was extracted from the FERET database [310], a collection of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles.

The frontal images in this database were extracted and from these, the images of subjects wearing glasses were manually selected and discarded since it has been shown that glasses make subjects more recognizable [311]. The resulting collection of images contained 3203 images of 1127 subjects.

A training set was then constructed from the resulting data set by selecting subjects with exactly three images. 203 such subjects (and hence, 609 images) were found. The remaining 2594 images of 924 subjects were declared as potential candidates for inclusion in probe and gallery sets.

The code for the chosen face recognition algorithms was acquired from the CSU face identification evaluation system [312]. The implementation details may be found in the user's guide for this system [313].

3.3.2.2. Preprocessing the data and training. The face images were normalized using the code provided in the CSU face identification evaluation system [312].

Training was conducted for each algorithm as described in section 3.2. The energy criteria proposed by Kirby [314] was selected for deciding the number of subspace basis vectors to retain and a threshold of 0.9 was chosen.

3.3.2.3. Computing the distances (similarity scores). A simple Euclidean distance measure was employed to compute similarity scores for the Eigenfaces and the Fisherfaces methods. Since Bolme et al [312] reported that the Mahalanobis Cosine measure [312] yielded better results for the Eigenfaces algorithm than the Euclidean one under some circumstances, therefore, this distance measure was also utilized for the Eigenfaces algorithm.

For the Bayesian intrapersonal/extrapersonal classifier, the similarity scores were calculated using both the maximum a posteriori (MAP) and maximum likelihood (ML) rules, as described in section 3.2.3.

3.3.2.4. Recognition criteria. A probe image was considered to be correctly recognized if a correct match for it was found within the first 1% closest matches returned by the algorithm.

3.3.2.5. Tests.

3.3.2.5.1. Large gallery test. The objective of this test was to examine the effect of increasing the number of subjects in the gallery on the recognition performance of the chosen techniques. Probe image sets of sizes 100, 200, 300, 400, 500, 600, 700, 800 and 900 were randomly selected from the available images (as described in section 3.3.2.1) such that no two images in one set were of the same subject. The corresponding gallery sets of the same sizes were constructed by selecting one image of each subject in the probe set such that that image was different from the one included in the probe set. The number of probe images in each probe set recognized by each method was recorded and a graph between the recognition rates and gallery sizes was constructed.

3.3.2.5.2. Image resolution test. The purpose of this test was to determine if changing the image resolution would affect the recognition rate of the selected techniques. The original resolution of the images was 256x384. The resolution of all the images (training as well as the gallery and probe ones) was reduced to 128x 192, 64x96, 32x48 and 16x24 respectively. The algorithms were retrained, the similarity scores were recomputed and the large gallery tests were carried out for each of these resolutions.

3.3.2.5.3 Number of bits per pixel test. This test was designed to study the effects of altering the number of bits per pixel of the face images on the recognition performance of the chosen methods. The original number of bits per pixel of the images was 8. This number was reduced to 4. The algorithms were retrained, the similarity scores were recomputed and the large gallery and image resolution tests were carried out again for this new number of bits per pixel.

3.3.2.5.4 Combinations of techniques. It appeared worth investigating if combining the outcomes of these approaches would produce better results than those generated by employing any of these methods individually. To this end, various combinations of these techniques were created and based on the findings of [315], the sum rule was applied to combine the normalized similarity scores of the individual methods to obtain the scores for the combinations.

The following combinations of techniques were examined:

C1. EF_Euc+EF_MahCos

C2. EF_Euc+FF_Euc

C3. EF_MahCos+FF_Euc

C4. Bay_MAP+Bay_ML

C5. EF_MahCos+Bay_MAP+Bay_ML

C6. EF_MahCos+Bay_MAP+Bay_ML+ FF_Euc

C7. EF_Euc+EF_MahCos+FF_Euc+Bay_MAP+Bay_ML

where EF=Eigenfaces, FF=Fisherfaces, Bay=Bayesian, Euc=Euclidean and MahCos=Mahalanobis Cosine.

All the tests, previously described, were conducted again for these combinations.

3.3.2.6. Results. The results of the tests for individual methods for different gallery sizes, resolutions and bits per pixel (as explained in the preceding section) are shown in Figure 3.1.

The following observations can be made based on these graphs:

- The recognition rates of all the methods decrease in general as the gallery size increases.
- The recognition rates of the Fisherfaces technique and both Eigenfaces methods remain stable (drop by only a couple of percentage points or so) as the resolution decreases to 64x96 (note that this confirms the results reported in [48]). However, the performance deteriorates rapidly after that. For the two Bayesian schemes, the recognition accuracy

actually improves in general as the resolution decreases to 32x48. However, it degrades as the resolution decreases further to 16x24.

- Changing the number of bits per pixel does produce some slight modifications in the results, causing the recognition rate to be slightly less at lower number of bits at some points. However, this difference is not really significant since these changes in the recognition rate, barring a few exceptions, are all well within a percentage point for the Bayesian schemes and within two percentage points for the Eigenfaces methods. The only exception is the Fisherfaces_Euclidean technique for which this difference is the most pronounced (approaches 4 percentage points), especially at lower resolutions. This indicates that this technique definitely does better with a higher number of bits per pixel.
- Here is a more detailed analysis of individual methods' performance with respect to change in resolution:
- The **Fisherfaces_Euclidean** method performs well at higher resolutions and produces the highest recognition rates in some cases for small gallery sizes. However, it suffers a sharp drop in recognition performance (40%-49% vs. 6%-25% for all other methods) at lower resolutions (i.e., 32x48 and 16x24) and it becomes the worst performing method at the lowest resolution of 16x24.
- **Eigenfaces_Euclidean** is obviously the worst performing at higher resolutions. Barring FF_Euc, it is the worst performing scheme at lower resolutions, too.
- The two **Bayesian** methods are the best performing ones in most cases at higher resolutions. In the cases where they do not produce the best recognition rates, their results are still comparable to the best performing method in those cases. They also perform the best in all cases at the low resolution of 32x48. Best performance results are equally

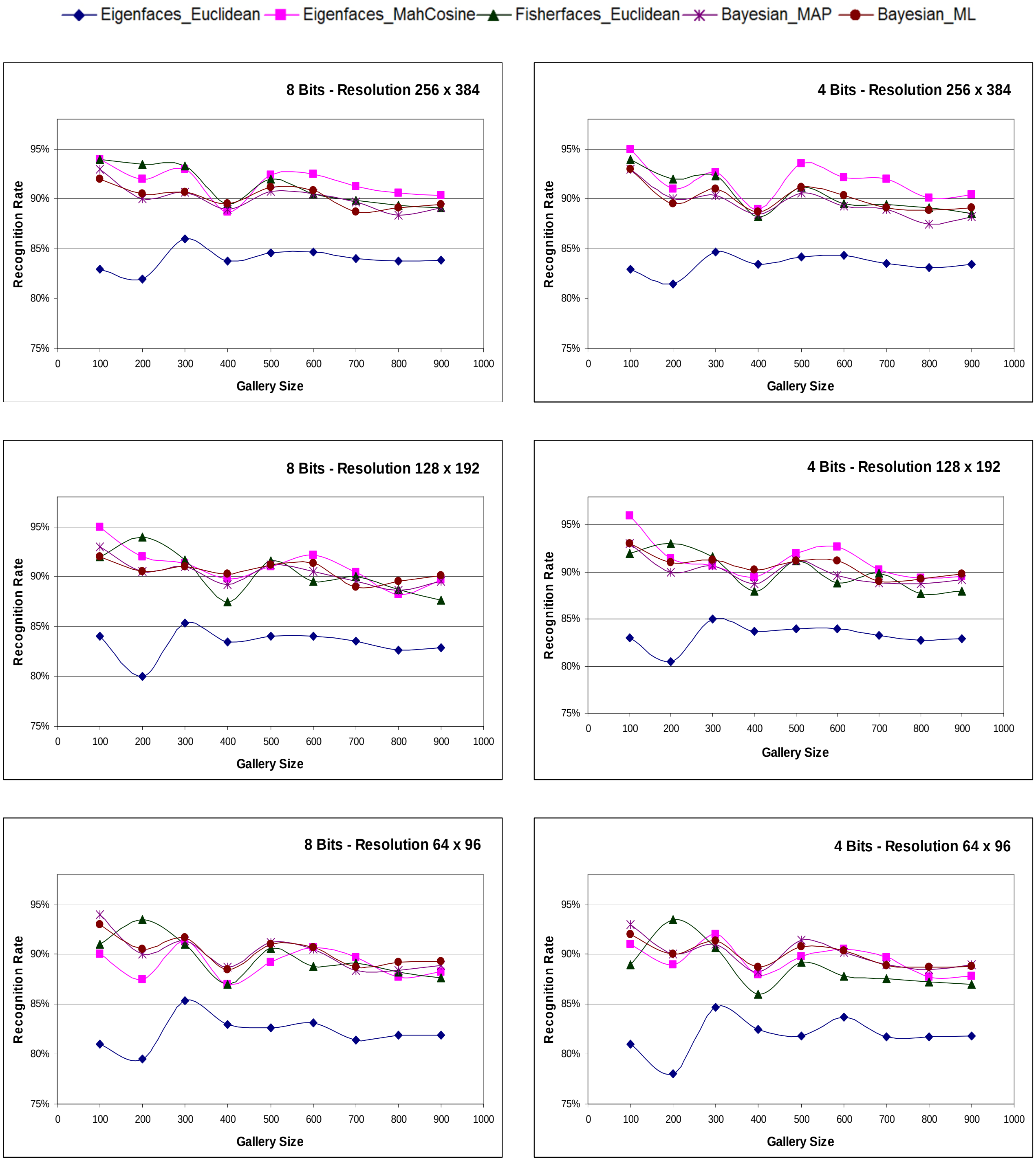


Figure 3.1. Results of tests for individual methods using various resolutions and numbers of bits per pixel (continued)

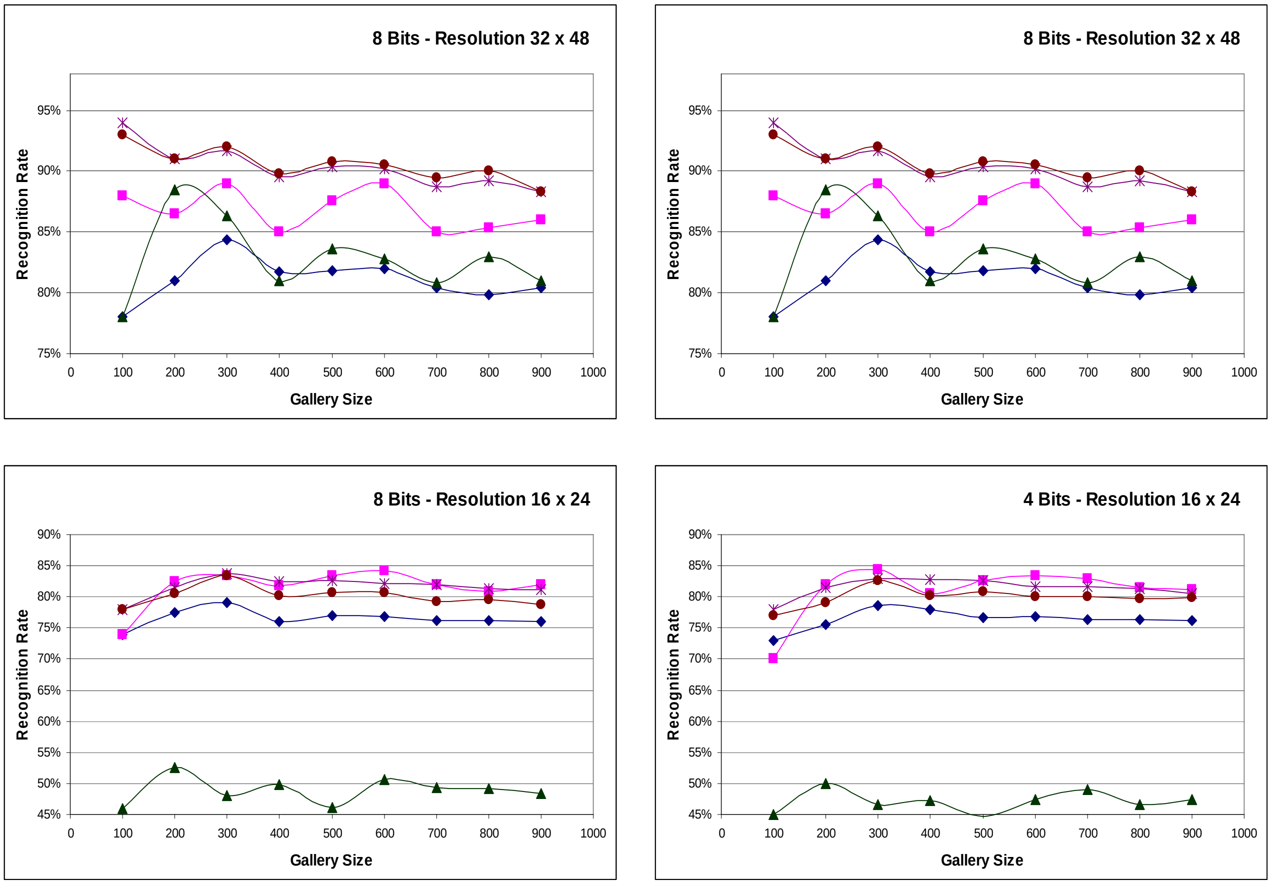


Figure 3.1. Results of tests for individual methods using various resolutions and numbers of bits per pixel

divided among Eigenfaces_MahCosine and the Bayesian techniques at the lowest resolution of 16x24.

- **Eigenfaces_MahCosine** is easily the best performing method at the highest resolution of 256x384 for larger gallery sizes. With the exception of resolution 32x48, it is the best performing or, at least, comparable to the best performing scheme at all resolutions, especially, for larger gallery sizes.

The results of the same tests for various combinations of techniques (as explained in the preceding section) are shown in Figure 3.2. The following observations can be made based on these graphs:

- The recognition rates of all combinations decrease in general as the gallery size increases.
- The 2nd observation for the Bayesian methods is true for the combination of these two methods too. The 2nd observation for the Eigenfaces and Fisherfaces techniques is true for the other combinations.
- In general, resolution changes cause recognition rates to drop the most for: a) the smallest gallery size for any combination, b) C2 (all gallery sizes) among all combinations (which is not surprising since this combination consists of the two methods whose performance drops the most at lower resolutions).
- Changing the number of bits per pixel to 4 does modify the performance especially for combinations containing the Fisherfaces algorithm. However, most of the differences in recognition rates are well within a percentage point and there does not appear to be a pattern suggesting that some method does consistently better with a certain number of bits per pixel. Therefore, the results using 4 bits per pixel have not been depicted here.
- C6 is consistently the best performing combination in the majority of the cases at all resolutions except the lowest one of 16x24. At that resolution, C5 is the best performing one (not surprising, since C5 is the same as C6 minus the Fisherfaces method and the Fisherfaces method's performance deteriorates significantly at the lowest resolution). C3 also seems to perform well for small gallery sizes at higher resolutions.

Here is a more detailed analysis of individual combinations' performances as compared to the methods comprising them with respect to change in resolution:

1. C1 performs better than EF_Euc in all cases. It does better than EF_MahCos in half the cases at higher resolutions and all cases at lower resolutions.
2. C2 performs better than EF_Euc but worse than FF_Euc in all cases for all but the lowest resolution. For the lowest resolution, it performs significantly worse than both EF_Euc and FF_Euc.
3. C3 performs better than the approaches comprising it at the two higher resolutions but performs worse than each of them at the three lower resolutions. Adding the Bayesian methods to this combination yields C7 whose performance is better than C3 in all cases. This suggests that the Bayesian methods do add some complementary information to that extracted by the EF_MahCos and FF_Euc methods.
4. C4's performance rarely supersedes that of both Bayesian techniques – its recognition accuracy lies somewhere between that of the two individual techniques. Since the recognition rates of the two Bayesian methods are very close to each other anyway, there does not appear to be an advantage to using this combination over the individual methods.
5. C5 (which combines the three best performing schemes) performs better than each of its constituent methods at higher resolutions but performs worse than them at the two lowest resolutions.
6. Adding FF_Euc to C5 yields C6 which produces yet better results than C5 for all but the lowest resolution. However, this combination's performance is still worse than that of EF_MahCos and FF_Euc at resolution 32x48 and worse than that of all the individual methods that it is comprised of at the lowest resolution of 16x24.
7. Adding EF_Euc to C6 (resulting in C7) improves the performance by an image or so in a

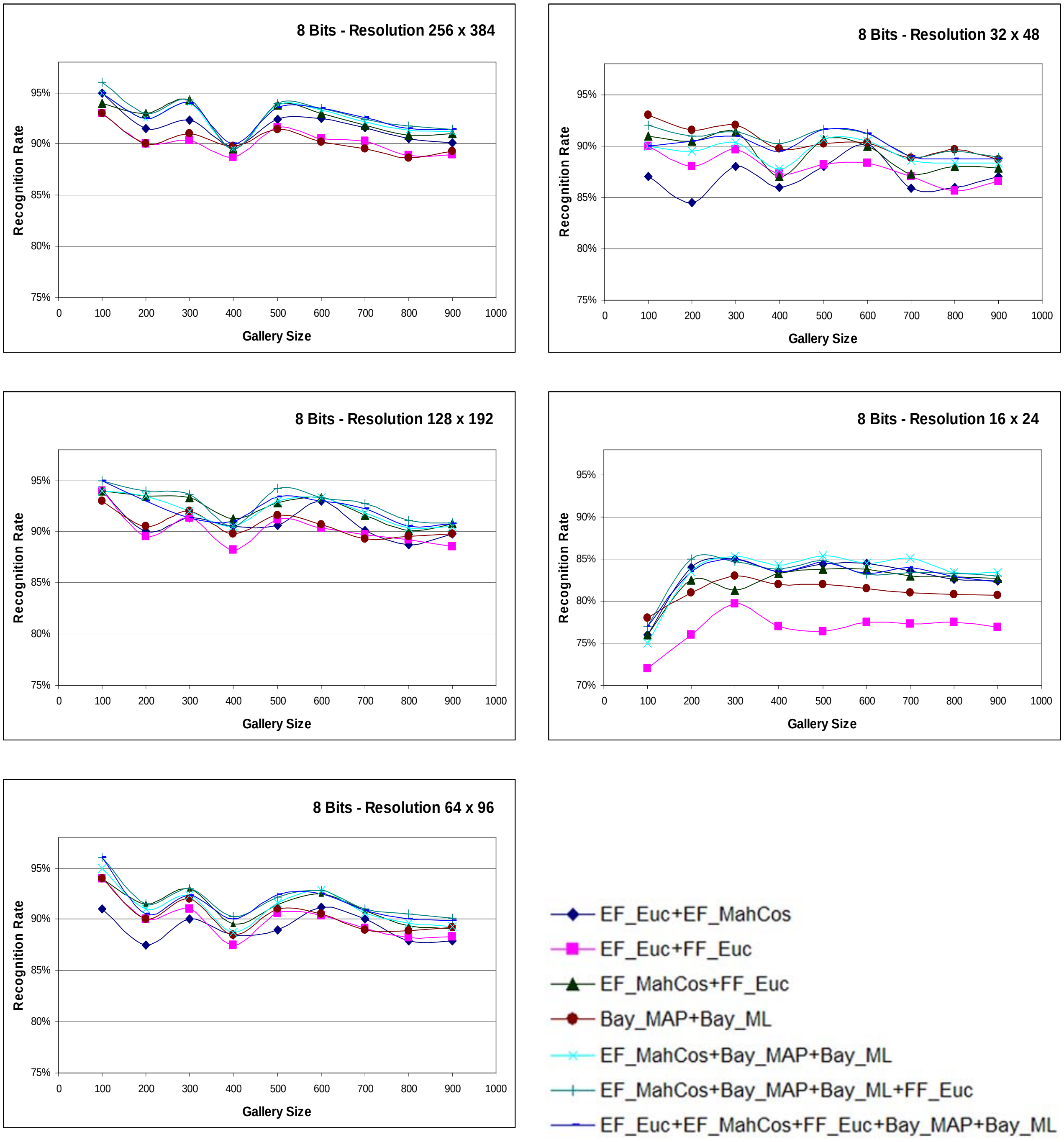


Figure 3.2. Results of tests for combinations of methods using various resolutions and number of bits per pixel

couple of cases and degrades it by an image or so in a few other cases. Thus, in general, the addition of this algorithm to C6 does not appear to offer any concrete advantage.

3.4. Conclusions

In this paper, some experiments to examine the performance of three algorithms, i.e., the Eigenfaces approach, the Fisherfaces technique, and the Bayesian-based Eigenfaces method, were presented. The recognition accuracy of these algorithms was tested for different gallery sizes, resolutions and bits per pixel. The same experiments were then repeated for various combinations of these techniques and the performance of these combinations with that of the individual methods was compared. This study has identified Fisherfaces_Euclidean as a method whose performance is significantly degraded by decreases in image resolution and bits per pixel. It has also shown that certain combinations of techniques perform better than the individual methods that those combinations are comprised of at certain resolutions. Even though the improvement in recognition accuracy is only up to 4 percentage points, but it is an improvement nevertheless warranting that it would be worthwhile to further investigate the unification of these approaches (maybe by examining if adjusting algorithm parameters (e.g., the number of eigenvectors retained) and using different training, gallery and probe sets further boosts the recognition rates).

CHAPTER 4

PCA-BASED METHODS FOR FACE RECOGNITION[2]

---

[2] Rabia Jafri and Hamid R. Arabnia. 2007. The 2007 International Conference on Security and Management (SAM'07), June 25-28, Las Vegas, USA. 534-541.
Reprinted here with permission of publisher.

Abstract

The statistical technique of Principal Components Analysis (PCA) has been utilized extensively for automated face recognition in the past several years and has been shown to perform well for this task. Several variations of this method have been proposed in recent years in terms of what kind of information is extracted from the face images as well as innovations in the method itself. An overview of these PCA-based approaches, with a focus on those which differ from the classical PCA-based face recognition method only in terms of the face image representation being used, is presented in this paper.

Keywords: Face Recognition, Person Identification, Principal Components Analysis, Face Image Representation.

4.1. Introduction

The well-known statistical technique of Principal Components Analysis (PCA) [316, 317] was first utilized for the task of automated face recognition by Turk and Pentland [6] nearly two decades ago. The encouraging results obtained from the preliminary experiments conducted by them fueled an abundance of research to explore the viability of this technique for the face recognition problem and a pattern of consistently favorable results reported since then have established this method as one of the most widely used ones for this task. The basic PCA-based technique has been used as a baseline for comparison of face recognition algorithms [305], been incorporated in commercial face identification products [318], and been the subject of extensive research spawning several variants which aim to improve its performance in terms of accuracy, speed and space requirements.

The popularity of PCA-based face recognition techniques stems from the advantages that they offer in terms of compactness of representation, high speed matching, ease of implementation, and non-stringent data requirements (i.e., they do not require full 3D models or detailed geometry). Furthermore, even bare-bones versions of these methods have been shown to perform well in relatively constrained conditions. Inspired by the promising results obtained by PCA-based methods, other statistical techniques have also been employed for face recognition (e.g., Fisher's Linear Discriminant [304], Independent Component Analysis [176]). However, most of these methods require a fairly large number of sample images per person in order to perform reasonably well. For the single sample image per person problem, which is the one most often encountered in real-life situations, especially, in security and surveillance applications, PCA-based methods are still considered to be among the strongest contenders [319].

In this paper, a review of the basic PCA-based face recognition method and some of its variants is provided. The paper is organized as follows: Section 4.2 briefly explains the basic PCA-based face recognition method proposed by Turk and Pentland [6]. Section 4.3 provides an overview of some variants of the PCA-based technique focusing on methods that differ from the original approach in terms of how the face images are represented. Section 4.4 concludes the paper.

4.2. PCA-based face recognition

An overview of the original PCA-based face recognition method proposed by Turk and Pentland [6] is provided below.

A face image of resolution NxN, can be thought of as a point in an N^2 dimensional space. Even if N is reasonably small, the dimensionality of this image space would be huge. However, since face images are so similar to each other in their overall configuration, they will form a tight cluster in this huge image space. The basic idea behind the PCA algorithm, when applied to the face recognition problem, is to find the vectors in this vast image space which best account for the variation among a given collection of face images. These vectors define the subspace in which the face images are clustered and are simply the eigenvectors of the eigensystem of the covariance matrix C of the given set of face images:

$$C = \frac{1}{M}\sum_{i=1}^{M}(x_i - \bar{x})(x_i - \bar{x})^T \qquad \textbf{(4.1)}$$

where M= number of images in the given set of images (called the training set), x_i = the *i*th image in the training set and $\bar{x}$ = the average of all the images in the training set.

Since the eigenvectors are linear combinations of the training images and are face-like in appearance, they have been termed as eigenfaces.

Note that C is an N^2xN^2 matrix, and solving its eigensytem is a computationally intractable task. However, it is possible to obtain the required eigenvectors by solving the eigensystem for a much smaller matrix of dimensions MxM and then taking linear combinations of the resulting vectors (See [6] for details). Since usually M<<N, therefore, the calculations henceforth become quite manageable.

The eigenvectors are now sorted in descending order of their associated eigenvalues and a certain number of those with the lowest eigenvalues are dropped (see [320] for common methods of deciding how many of the lower-valued eigenvectors to eliminate). One obvious reason for not using all the eigenvectors is computational efficiency but an even more important one is generalization. A recognizer that uses all the information in the training set is inevitably prone to using some random correlations specific to that set to produce better classification for only that set and does not generalize well to other data sets [321]. By selecting the largest-variance eigenvectors that encode the most useful information, and discarding the lowest-variance eigenvectors, which, it is generally agreed, mainly encode noise, the generalization capabilities of the classifier are greatly increased.

The images in the database of known individuals are then projected onto the remaining M' ($<M$) eigenvectors to form face class vectors which are stored as compressed representations of the known face images. The probe image to be recognized is also projected onto the M' eigenvectors and the similarity scores between the resulting projection vector and each of the face class vectors are computed using a similarity criteria or distance measure of choice. The probe image is then declared to be that of the individual whose face class vector is most similar to the projection vector of the probe image (provided that the similarity score is above some threshold theta; if not, then the probe image is classified as unknown).

The method was tested using a database of 2500 images of 16 people under all combinations of 3 head orientations, 3 head sizes or scales and 3 lighting conditions. Though the technique appeared to be fairly robust to lighting variations, and less so to changes in orientation, but its performance was found to degrade significantly with scale changes. Later, several other experiments were also conducted on a much larger and diverse database (containing 7562 images of 3000 individuals) to evaluate the recognition and verification accuracy of this approach as well as to test its sensitivity to racial differences [322]. The recognition results obtained were encouraging in general and the verification accuracy was found to be comparable to those of other biometric techniques.

4.3. Variants of the original PCA-based face recognition approach

The original PCA-based recognition method discussed in the previous section suffers from several drawbacks, a few of which are outlined here: it requires a high degree of correlation between the test and training images and does not perform well under large variations in pose, scale and illumination etc. Furthermore, PCA effectively sees only the Euclidean structure and fails to discover the underlying structure if face images lie on a non-linear submanifold in the image space (which is the case especially if there is perceivable variation in viewpoint, illumination or facial expression) [167]. Also, PCA considers only the pairwise relationships between pixels in the image database and fails to take the higher order correlations among them (which may be crucial to better represent complex patterns) into account [323]. Last, but not least, PCA is an unsupervised learning technique which does not take class information into consideration and thus, when it maximizes the total scatter across all projected samples, it not only maximizes the between-class scatter, which is useful for discrimination, but also the within-class scatter, which confounds the classification results [304].

To overcome these drawbacks, several alternatives to the eigenfaces approach have been proposed (e.g., techniques to discover the nonlinear structure of the manifold in the image space where the face images lie, such as Isometric Feature Mapping (ISOMAP) [167] and Laplacian Eigenmap [169]; methods that do take class information into account, such as those based on Linear Discriminant Analysis [304] [142, 324], etc.). However, most of these alternative approaches fail when only a single sample image per person is present because of the absence of intra-class variations and, in case of nonlinear techniques, have high computation times relative to those of the eigenfaces method. For the single sample image per person problem, PCA-based methods have repeatedly been shown to outperform other subspace-based face recognition techniques [319].

Therefore, despite the emergence of several alternative face recognition methods, much effort has been devoted and extensive research continues to be conducted to modify/extend the eigenfaces approach itself to enhance it in various ways:

- To deal with pose variations, Pentland et al. [325] generalized the eigenfaces approach, which had been tested only on frontal human faces, to other views. Moreover, they introduced the concept of modular recognition where prominent facial features (such as the eyes, the nose and the mouth) are extracted from each face image and these features are used either by themselves or in conjunction with the whole face image for recognition using the same process as described above. A variation of their technique was later implemented by Kumar et al. [26] who, instead of extracting distinct facial features, divided the face image into equal-sized subimages. It should be noted that Pentland et al.'s idea of dividing the face image into regions, applying a subspace-based face recognition method to each region and combining the outcomes has been the focus of much research in recent years and has yielded promising results in general (please see [45] for a review).
- To overcome the illumination variation problem, Bischof et al. [326], Epstein et al. [327], Hallinan [328], Ramamoorthi [329], and Zhao and Yang [330] have analyzed the ways of modelling the arbitrary illumination condition for PCA-based recognition methods.
- Generalizations of PCA, which do address higher-order statistical dependencies in the image set, have also been utilized for the face recognition task. Examples include

methods based on Independent Component Analysis (ICA) [176, 177, 182] and Kernel PCA [323, 331, 332].

- To overcome the inability of the PCA method to take class information into account, some variants have been presented which incorporate face class labels in the classification process. Examples include Class-Augmented PCA [333] and class-information-incorporated PCA [334].
- Substantial work has also been done to explore if using different similarity criteria and distance measures significantly alters the recognition rates obtained by the eigenfaces method [309, 320, 335].
- Several methods have been proposed in which PCA is used to reduce the dimensionality of the image vectors and then some other technique is applied to classify the reduced vectors. Examples of techniques used in combination with PCA include Fisher's Linear Discriminant [304], Support Vector Machines [336], neural networks [321] and canonical correlation analysis (CCA) [337].
- Speeding up the recognition process when large databases of faces are involved has been the focus of other research where the computation time of a PCA-based method is reduced by modifying the method itself as well as by utilizing various parallel architectures [338-340].

4.3.1 Variants of eigenfaces in terms of the face image representation used

It is evident from the above discussion that many variants of the original PCA-based face recognition method have been proposed in recent years. Many of these approaches modify the PCA technique itself or alter the way in which it is applied. However, several researchers have observed that the vector representation of the face images which is utilized in the eigenfaces method might not be the best for classification purposes and have, therefore, suggested that the face images be transformed to a new representation before applying PCA. These techniques are, thus, the same as the eigenfaces method except for the face image representation being used. In this section, we shall focus on this class of variants of the original PCA-based approach. A brief overview of some such methods is given below.

4.3.1.1 Two-dimensional PCA

In the eigenfaces method, the image matrices are converted into one-dimensional vectors and those vector representations of the images are used in all subsequent computations. The two-dimensional PCA technique [341] differs from eigenfaces only in that the image matrices are not converted into one-dimensional vectors but are rather used directly in all computations. The image covariance matrix is calculated the same way as in equation (1). However, since the matrix representation of the images is now being used, the resulting covariance matrix is of dimension NxN rather than N^2xN^2. Evaluating this significantly smaller matrix accurately is much easier and calculating the corresponding eigenvectors is much faster relative to the eigenfaces method.

However, projecting an image onto the eigenvectors in this case will yield a set of vectors which are then used to form a matrix called the *feature matrix* of that image. Note that in the eigenfaces method, projecting the image onto the eigenvectors yields a set of scalars, which are then used to form a *feature vector* of the image. Since the *feature vector* of PCA is smaller than the *feature matrix* of 2D-PCA, hence, 2D-PCA is at a disadvantage relative to the eigenfaces approach in terms of storage requirements as it requires more coefficients to represent an image.

The algorithm was evaluated using the following three face image databases (the factors tested for each database are given in parentheses): 1)ORL [342] (pose, sample size), 2)ARL [343] (time, expression, illumination), 3)Yale [344] (expression, illumination). The recognition accuracy of 2D-PCA was found to be better than eigenfaces in all the experiments though, in some cases, the difference in performance was not statistically significant. 2D-PCA was also found to be computationally much more efficient than the eigenfaces method in all the tests. For the ORL and Yale databases, the recognition accuracy of this method was also compared to other techniques, i.e., Fisherfaces [304], ICA [177, 345], and Kernel Eigenfaces [346], and was found to be better in all cases (except than Fisherfaces in the "leave-one-out" strategy experiments on the ORL database).

4.3.1.1.1 Variants of two-dimensional PCA

Zhang et al. [347] showed that 2D-PCA essentially works in the row direction of images and proposed an alternative approach, called two-directional 2D-PCA, which simultaneously considers both the row and column directions. Experiments carried out on the ORL [342] and FERET [32] databases showed that this method attains the same or even higher accuracy than the 2D-PCA approach and requires a much reduced coefficient set for image representation.

Kong et al. [348] also proposed a technique to reduce the coefficient set required by 2D-PCA for representing images. Their method, called Bilateral-projection-based 2DPCA (B2DPCA), achieves this reduction by projecting the row and column vectors of the image matrices to two different subspaces. They also pointed out that though 2D-PCA has been shown to perform better than the standard PCA method, however, it is still a linear projection technique which is incapable of taking higher order relationships among the row/column vectors into account. They, therefore, have presented a Kernel based 2D-PCA (K2DPCA) approach to remedy this drawback. Tests conducted on the ORL [342], UMIST [349] and Yale [344] databases showed that B2DPCA and K2DPCA yield better recognition rates than the 2D-PCA method.

4.3.1.2 Diagonal PCA

Zhang et al. [21] argue that the 2D-PCA method reflects only the variations among the image rows and not the variations among the image columns, which are also important for recognition. They attempt to remedy this shortcoming by proposing the Diagonal PCA (DiaPCA) technique which converts image matrices into diagonal face images and then utilizes those diagonal images for recognition following the same process as 2D-PCA. The rows (columns) in the diagonal images simultaneously integrate row and column information and, therefore, reflect correlations between both row variations and column variations. Let A be an image of resolution mxn. Figures 4.1(a) and 4.1(b) show how the diagonal image B is extracted from A if $m \leq n$ and if $m>n$ respectively. They also propose combining Diagonal PCA and 2D-PCA (DiaPCA+2D-PCA) to examine if that would further enhance the recognition accuracy.

The proposed methods were evaluated by conducting tests on a subset of frontal images, containing expression variations, from the FERET database [32]. The top recognition accuracy of DiaPCA was found to be greater than that of 2D-PCA, while DiaPCA+2D-PCA performed yet better than DiaPCA. Moreover, the running times of 2D-PCA, DiaPCA and DiaPCA+2D-PCA were found to be comparable and significantly less than those of the standard PCA-based approach (which is not surprising since the covariance matrices of these methods are much smaller than the standard PCA technique and calculating the eigenvectors of these matrices, therefore, takes considerably less time).

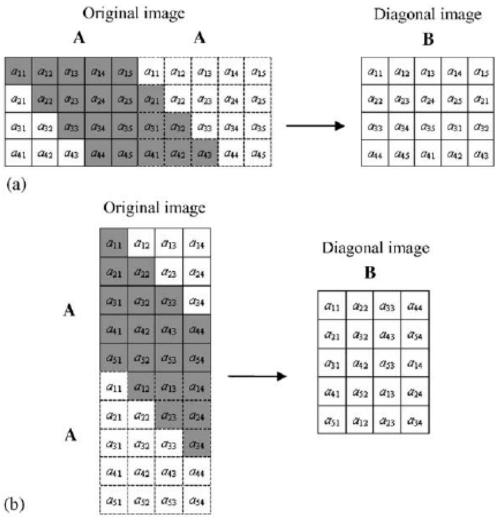


Figure 4.1. Illustrations of the ways of deriving the diagonal face images [21]. (Reprinted with permission from Elsevier)

4.3.1.3. Symbolic PCA

The eigenfaces approach considers each pixel in an image as a feature and represents it by a single quantitative value. In the Symbolic PCA [350] method, however, each face class is represented as a symbolic object and each feature can then be viewed as a qualitative multi-valued symbolic variable. In this approach, a symbolic face is first constructed for each face class as follows: Let the training data consist of face image vectors of length p of m subjects, with an equal number of images per subject (the images may contain variations in expression, orientation, and illumination). Then the feature vector for a subject is described by a vector of p interval variables $Y=(Y_1, \ldots, Y_p)$, where $Y_j=[\min(x_j), \max(x_j)]$, $j=1,\ldots, p$. $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum intensity values, respectively, among the jth pixels of all the images of that subject. The vector Y is the symbolic face for the subject. The symbolic face vector Y is then converted into a vector of scalars X by using the centers method [351]: $X=(X_1,\ldots, X_p)$, where $X_j=(\min(x_j)+\max(x_j))/2)$, $j=1,\ldots, p$. These face class vectors are the used to calculate the covariance matrix as in equation (1). The rest of the training and the classification process is the same as the eigenfaces method.

The performance of this technique was tested on the ORL database [342] using four different similarity measures, namely, City Block (L1), Euclidean (L2), Mahalanobis and cosine measures. The best results were obtained using the cosine measure and thus, that measure was used in all subsequent experiments. A comparison of Symbolic PCA (using the cosine measure) with eigenfaces (using the Euclidean measure) showed that Symbolic PCA performs better than eigenfaces both in terms of recognition accuracy as well as computation cost.

4.3.1.4. Line-based PCA

In the line-based PCA approach [22], a face image is first divided into equal sized blocks. Next, the face image is transformed into a column vector of these blocks. Each block is then converted into a vector by concatenating its rows or columns. Hence, the face image vector now becomes a two dimensional matrix with each row of that matrix representing a block of the original image. The process of transforming a face image into this new representation is illustrated in Figure 4.2.

The recognition process is the same as the 2D-PCA method except that, instead of the original 2D face image matrices, the new 2D representations of the face images are used.

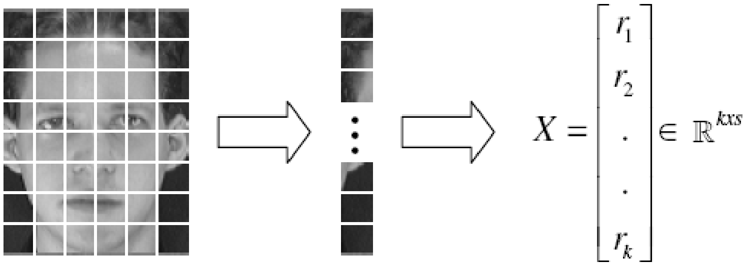


Figure 4.2. The process of transforming the original face image to its new representation [22]. (With kind permission of Springer Science and Business Media)

Experiments were conducted on the ORL database [342] using different numbers of training images and a block size of 3x3 pixels. The top recognition accuracy of the proposed method was found to be better than that of the eigenfaces technique in all cases. Further tests carried out to determine the optimum block size (block sizes of 2x2, 3x3, 5x5, 10x2, 10x10 pixels were evaluated) showed that the block size of 5x5 pixels yielded the best recognition rates.

4.3.1.5. Summary of variants of eigenfaces in terms of the face image representation used

A summary of the approaches discussed in this section, which vary from the original PCA-based face recognition method (i.e., eigenfaces) in terms of the face image representation used, is given in Table 4.1.

All these methods have been reported to perform better than the eigenfaces technique. It should be noted that an objective comparison of these techniques is not practical since the experiments conducted to evaluate these methods differ from each other in terms of the face databases utilized, the number of images used, the factors tested, and PCA parameters (e.g., the number of eigenfaces used). It would be interesting to compare the performance of these methods on the same dataset under the same experimental conditions to determine which of these face representations yields the highest recognition rates.

Table 4.1. Summary of variants of eigenfaces in terms of the face image representation used

Approach	Face image representation	Face databases used	Factors tested
2D-PCA [341]	2D matrices	ORL [342], ARL [349], Yale [344]	Pose, sample size, time, expression, illumination
Diagonal PCA [21]	Diagonal face images in the form of 2D matrices	FERET [32]	Expression
Symbolic PCA [350]	Centered symbolic face vectors for training images, 2D matrices for probe images	ORL [342]	Scale, illumination, expression, time
Line-based PCA [22]	2D matrices with each row representing a block of the original image	ORL [342]	Scale, illumination, expression, time

4.4. Conclusions

The PCA-based techniques presented in this paper have improved upon the standard eigenfaces method in several ways. Some methods have enhanced the recognition accuracy while others have reduced the computations, space and time required. The encouraging results obtained so far, coupled with the fact that PCA-based algorithms still appear to be the strongest contenders in the field of face recognition for scenarios where only a very small number of images per person are available, definitely warrant further research in this area.

CHAPTER 5

A SURVEY OF COMPONENT-BASED FACE RECOGNITION APPROACHES[3]

---

[3] Rabia Jafri and Hamid R. Arabnia. 2007. The 2007 International Conference on Artificial Intelligence (ICAI'07), June 25-28, Las Vegas, USA. 103-113.
Reprinted here with permission of publisher.

Abstract

Currently, most of the state-of-the-art face recognition methods utilize the image of the whole face to perform classification. However, in recent years, several techniques have been proposed which segment the face image into various parts, apply known face recognition methods to those facial regions and finally, combine the recognition results thus obtained to determine the identity of the whole face. A survey of some of the most widely known methods in this category is presented in this paper and a detailed analysis of their performance as well as a discussion of how they compare to approaches which utilize the whole face without subdividing it is provided.

Keywords: Face recognition, Person identification, Facial regions.

5.1. Introduction

Automated face recognition is presently an important problem in the field of computer vision due to its myriad applications in areas such as surveillance, security and human-computer interaction [50, 51, 55]. Numerous approaches have been proposed to solve this problem over the past few decades (please see [352] for a detailed survey) and even though prevalent state-of-the-art systems for this task do perform reasonably well under constrained circumstances, however, the general problem of identifying faces under real-life conditions involving variations in pose, illumination, expression, age and occlusions/distortions remains largely unsolved to date.

Until recently, face recognition techniques operating on intensity images could be broadly divided into two categories: 1) Feature-based methods, which first process the input image to identify and extract distinctive facial features such as the eyes, mouth, nose, etc. as well as other fiducial marks and then compute the geometric relationships among those facial points, thus, reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed for matching faces using these measurements. (For examples, please see [101, 104]). 2) Appearance-based (or holistic) methods, which attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face (for examples, please see [6, 8, 176]). However, in the past few years, a new group of techniques, generally referred to as component-based methods, has emerged whose aim is to combine the advantages offered by both these classes while offsetting/minimizing their drawbacks. These approaches first divide the face image into various regions or components and then classify those components rather than the whole image alone. The classification results are then combined to determine the identity of the individual whose face appears in the image. Several such techniques have been proposed in recent years and, since they have reportedly produced promising results overall, it appears worthwhile to conduct a survey of such methods to get a general idea of their particulars (i.e., which algorithms are utilized, how specifically is the face segmented, what kind of tests have been conducted to evaluate their performance, etc.) as well as a feel for how such approaches fare as compared to those which consider the whole face image without subdividing it.

The rest of this paper is organized as follows: Section 5.2 outlines some of the advantages that component-based methods appear to offer over feature-based and appearance-based techniques. Section 5.3 contains a review of some well-known recent approaches belonging to this class of techniques and an analysis/discussion of how they compare with methods which consider the whole face as one unit. Section 5.4 contains some concluding remarks.

5.2. Motivation for using a component-based approach

Both feature-based and appearance-based approaches come with their own set of pros and cons. The main advantage offered by feature-based methods is that, since the extraction of the feature points precedes the analysis done for matching the image to that of a known individual, in principle, such techniques can be made invariant to size, orientation and/or lighting [101]. Other benefits of these methods include compactness of representation of the face images and high speed matching [116]. The major disadvantage of these approaches is the difficulty of automatic feature detection and the fact that the implementer of any of these techniques has to make arbitrary decisions about which features are important [117]. After all, if the feature set lacks discrimination ability, no amount of subsequent processing can compensate for that intrinsic deficiency [101].

Appearance-based or holistic methods' major strength lies in their not destroying any of the information in the face images by concentrating on only limited regions or points of interest [83]. However, this same property is also their greatest drawback, since most of these approaches start out with the basic assumption that all pixels in the image are equally important [117]. Consequently, these techniques are not only computationally expensive but require a high degree of correlation between the test and training images and do not perform effectively under large variations in pose, scale and illumination etc. [17]. Though various dimensionality reduction techniques have been exploited to represent the image data more compactly by retaining only the most meaningful dimensions, however, such techniques may also potentially cause the loss of discriminative information [188].

Component-based approaches strive to combine the benefits of both these groups while reducing the drawbacks associated with them. Since these techniques consider facial regions rather than specific points in an image, they have access to more data (and, arguably, more discriminatory information) as compared to the feature-based methods. Even though they utilize the same algorithms that are employed by appearance-based approaches, however, since the statistical complexity of a facial component is less than that of the whole face image, therefore, applying a linear encoding via statistical techniques such as PCA/LDA/ICA to a facial region, rather than the whole face, is likely to result in better classification [159]. Also, pre-processing of small regions is easier in general than that of the whole image [353]. Furthermore, variations in expression, illumination and pose acutely affect some facial regions while leaving other segments of the face unchanged. The unchanged components will closely match their corresponding components in an image taken under normal conditions. Even for the affected regions, changes due to slight variations in pose or lighting will be smaller compared to the changes in the whole image pattern [30]. The same argument can be made for partially occluded or decorated (e.g., bearded or bespectacled) faces. Therefore, component-based methods should perform better than their holistic counterparts in such scenarios. Moreover, Kim et al. [353] argue that though variations in lighting and orientation would cause non-linear changes in the image intensity function for the whole face, however, in the face components, these effects can be approximated by a linear function, which can be corrected for much more easily. The reasons outlined above do appear to provide ample justification for a component-based approach for face recognition. We shall now proceed to examine some such techniques introduced in the past few years to determine if such methods do indeed work better in practice.

5.3. Review and analysis of component-based face recognition systems

Brunelli and Poggio [2] were among the first to implement a component-based system for recognizing faces. In their approach, the eyes are first detected (through correlation matching eye prototypes) in all face images and the images are then normalized in terms of scale and orientation. Each person is represented by a set of four masks: eyes (including eyebrows), nose, mouth and face as shown in Figure 5.1. The location of these features is relative to the eye positions. A new face is compared to that of a known person by computing similarity scores for all 4 features using normalized cross-correlations. The unknown person is then classified as the one giving the highest cumulative score. Since correlation is sensitive to illumination gradient, 4 different normalization techniques were tried and the one using gradient magnitude information was found to be the best. The system was tested on a database of 47 people (26 male and 21 female) with 4 images (frontal, 512x512 pixels, with constant illumination) per person. The results indicate that this method had better performance as compared to a feature-based approach (which utilized 35 numerical features such as the vertical position and width of the nose and mouth and the bigonial and zygomatic breath of the face) on the same data.

Figure 5.1. The different regions used in the template matching strategy [2] (©1993 IEEE)

Another pioneering effort in this domain is that of Pentland et al. [1], who introduced a modular eigenspace approach, an extension of the PCA-based eigenfaces technique proposed in [6]. In their method, facial features (i.e., the eyes, the nose, and the mouth) are first detected using the same process which is used for detecting faces in [6]. An eigenspace ("feature space") for each feature is then constructed using sample features extracted from a representative set of face images (the training set). The similarity score between two images can then be computed as follows: for each image, the features are extracted and projected onto their corresponding feature spaces; the Euclidean distance between each pair of corresponding projected feature vectors (eigenfeatures) is then computed; finally, these distances are summed up to get the similarity score between the two images. Given a face image of an unknown person, the similarity scores between that image and the images of all known individuals are computed and the person is classified as the one with the highest similarity score. Tests were conducted on a database of 45 people with two images per person, one smiling and one neutral. The training set consisted of the neutral images while the test set was comprised of the smiling ones. The facial features used

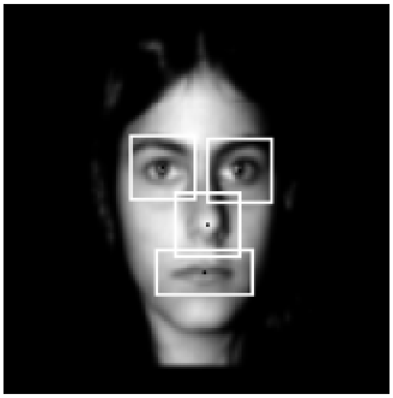


Figure 5.2. Facial components used in the modular eigenspaces method [1] (© 1994 IEEE)

were the two eyes and the nose. It was found that the eigenfeatures performed as well as the eigenfaces method achieving a recognition rate of 95% when using 20 eigenvectors. Furthermore, in the lower dimensions of eigenspace, eigenfeatures yielded significantly better results than the eigenfaces approach. Combining the similarity scores of the eigenfeatures and eigenspaces improved the recognition rate further (up to 98%). Some additional tests were also conducted where some test images of the same individual with gross variations (e.g., a hand on the cheek, a painted face, a beard) were introduced. The eigenfeatures method was able to find the correct matches for these images, whereas the eigenfaces method failed to do so.

The encouraging results reported by Pentland et al. [1] spawned off an abundance of research in this area. Most of the component-based approaches introduced since then have been modeled on their method in that they use a subspace projection-based face recognition system and the general recognition process is the same as the one described above (i.e., facial components are extracted, training is conducted to find the subspace for each component, facial components of known and unknown subjects are projected into their corresponding subspaces, similarity scores are found by computing the distances between the projections, some rule is used for combining the similarity scores of all components, and finally, the unknown person is classified as the one giving the highest combined similarity score). The differences among these techniques lie in the preprocessing of the face images, the selection of the facial components, the subspace projection methods employed and the rules for combining the similarity scores. Since these methods form the bulk of component-based techniques for face recognition, the rest of this section will outline several of these approaches, followed by a brief discussion of techniques which do not conform to this model.

5.3.1. Component-based methods modeled on the modular eigenspaces method

Brief descriptions of some techniques modeled on Pentland et al.'s approach [1] are provided in this section.

Price and Gee's method [23] utilizes the following three components (shown in Figure 5.3(a)): the whole face, the eyes and nose region, the eyes only region (which includes the eyebrows). The images are preprocessed (geometrical normalization based on labeled landmarks, application of elliptical mask to remove background, normalization to have zero-mean and unit variance). The three components are then manually extracted from each image. For dimensionality reduction, a new algorithm, called direct, weighted LDA (DW-LDA), has been introduced which is a combination of the direct LDA (D-LDA) [137] and the weighted pairwise Fisher criteria (W-LDA) [143]. The other algorithms utilized for dimensionality reduction are: PCA, PCA/LDA, PCA/W-LDA, and D-LDA. The face images for evaluating the system were acquired from the CVL [354] (114 subjects with 3 frontal images per subject (neutral, smiling without teeth showing, smiling with teeth showing)) and Yale [344] (15 subjects with 11 images per person (variations in expression, illumination, and decoration)) databases. Some of the images were discarded for various reasons so that the resulting database consisted of 484 images of 128 subjects. 100 training and classification runs were performed. In each run, one randomly selected image of each subject was used for training while the rest were utilized for testing. The spaces for the components were reduced to 50 dimensions for all algorithms. The results indicate that PCA/W-LDA is the best performing subspace method and also that the eyes-only component is better for classification than the other two components.

In Toygar and Acan's approach [24], preprocessing involves cropping out only the face part (from the forehead to the chin) from each image. The resulting image is then divided into 5 equal-width horizontal segments as shown in Figure 5.3(b) (some initial tests were carried out using 3, 5, 7 and 9 segments and using 5 segments appeared to yield the best results). The dimensionality reduction methods employed are PCA, LDA and ICA. Also, several different methods are used to combine the similarity scores, namely, majority voting [355], Borda count, highest rank method [356], sum rule, max rule and the median rule [315]. Tests were conducted on frontal images of 50 subjects (2 samples per person, scaled down to 45x35 pixels) from the FERET database [32]. In the first group of tests, the performance of individual component classifiers was compared against the holistic classifiers. It was found that holistic PCA did better, holistic LDA did slightly worse and holistic ICA did significantly worse than each of their corresponding component classifiers. In the second set of experiments, the component classifier scores were combined using the combination rules described above. The sum rule and the Borda count method were found to be the best and the performance of the combined classifier using the sum rule was then compared to that of the corresponding holistic algorithms. This time, holistic PCA did as well as but not better than, holistic LDA did significantly worse than and holistic ICA did much worse (yielding 40-50% less recognition rates) than its corresponding combined classifier. In the third group of tests, the probe set consisted of 300 images. Each of these images varied from its corresponding gallery images in terms of one or more of these factors: time, orientation, illumination and scale. In this case, holistic PCA did slightly better than, holistic LDA did worse than and holistic ICA did much worse than its corresponding combined classifier. It is also shown that the combined classifier approaches have significantly lower time and space requirements than the holistic methods.

Yan and Osadciw [25] preprocess the images by cropping out the face, downsampling the cropped images to a common size and adjusting the illumination and the head leaning. 4

components are then extracted from each image: the eyes, the nose, the mouth and the forehead. The whole face is considered as the 5th component. PCA is used for dimensionality reduction in this case. Two methods are used to combine the scores of the individual component classifiers: score level fusion (the similarity scores of the component classifiers are summed up; each component classifier may be assigned a different weight) and decision level fusion (the best matches of each of the component classifiers are fed to a majority voting machine which makes the final decision). Experiments were performed on frontal images taken from the ORL [342], AR [343], Yale [344], UMIST [349] and FERET [32] databases, so that there were variations in lighting, expression and head leaning in the test data. Both combination rules yielded better results than any of the individual classifiers. Score level fusion, with equal weights for all components, performed better than decision level fusion and produced the best results. However, it is pointed out that decision level fusion is simpler in terms of data transmission and implementation.

Kumar et al. [26] divide the face into 3 equal-width horizontal segments as shown in Figure 5.3(d). Their method differs from Pentland et al.'s [1] in that the similarity score for each segment is multiplied by a weight which represents "a measure of the extent of variation in eigenspace for a sub-region of a subject across all samples." The final similarity score is found by adding the weighted similarity scores for all segments. Their method was tested using images from the Yale database [344] which contain expression and illumination variations. The test results showed that their technique performed significantly better than both conventional PCA and Pentland et al.'s method [1].

Zhang et al. [27] crop out the face using a mask and scale each image down to 32x32 pixels. The face is divided into 4 overlapping components as shown in Figure 5.3(e). PCA, followed by LDA, is the subspace method utilized for each component classifier. However, their approach differs from the ones mentioned so far in that the projection vectors of all the components are combined into a single vector and then LDA is applied to this new combined vector to get the final face descriptor. The system was tested using frontal images differing in expression from the FERET database [32]. The training set, gallery and probe set consisted of 1002, 1196 and 1195 images respectively. Their method yielded slightly better recognition rates as compared to conventional LDA (93.47% versus 91.8% at rank 1).

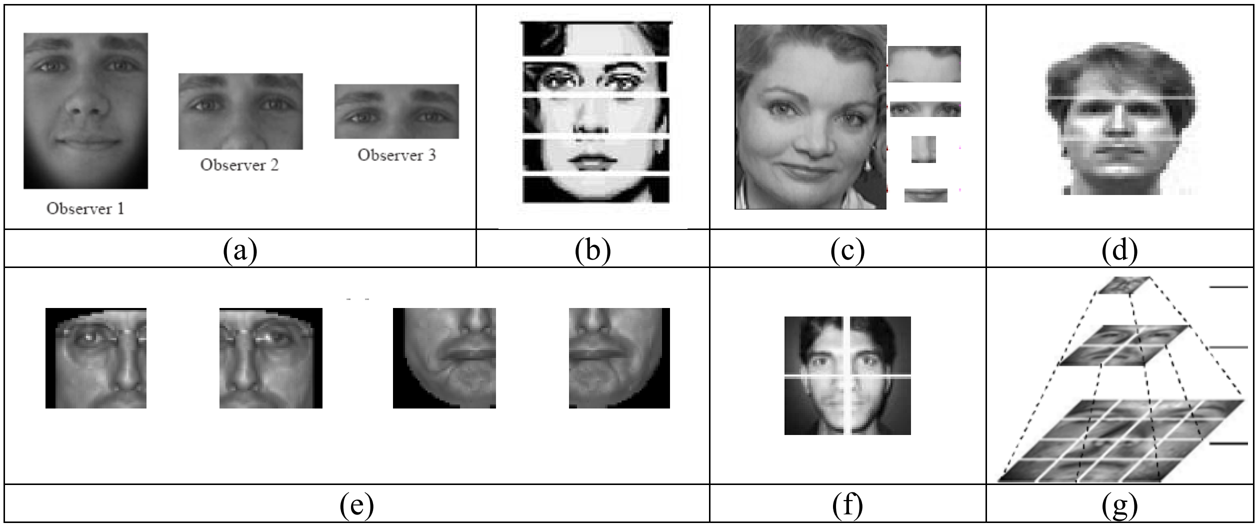
Gottumukkal and Asari [28] normalize and crop the face images to a size of 64x64 pixels. Their method of obtaining facial components differs from the ones previously mentioned in that they do not extract facial regions based on the detection of salient facial landmarks such as the eyes, nose etc.. Rather, they simple divide the face image into *N* subimages of equal size. Dimensionality reduction is done using PCA. The similarity score for a probe image is computed by taking the average of the similarity scores of all its subimages. To test the performance of this technique for large pose variations, 10 images each of 20 people were taken from the UMIST database [349]. 8 images were used for training and 2 for testing. The number of subimages was varied from 4 to 4096. The results indicate that their approach did not yield significantly better results than the conventional PCA one in this scenario. To evaluate this scheme for large variations in expression and illumination, 11 images each of 15 people were taken from the Yale database [344]. 8 images were used for training and 3 for testing. Again, the number of subimages was varied from 4 to 4096. In this case, the results showed that their approach completely outperformed the holistic PCA one for *N*= 4, 16, 64 (best results were for *N*=16). Some additional tests were performed on the Yale database using a leave-one-out strategy.

Results for N=16 indicate that their approach outperforms conventional PCA. They also claim that their method performs better than Pentland et al's modular eigenspaces approach [1].

Su et al. [29] also crop the face images to a size of 64x64 pixels and divide each face image into N subimages of equal size. However, instead of working with the pixel intensity values, as all the methods previously mentioned do, their technique extracts the Gabor features in each subimage and the dimensionality reduction method - FDA in this case - is then applied to those Gabor features. The sum rule is used to combine the similarity scores from all subimages. They assert that this approach results in more accurate classification than the Gabor Fisher Classifier (GFC) [357] and Adaboosted Gabor Fisher Classifier (AGFC) [325] techniques (which operate similarly but work with the whole image rather than subimages, and, thus, tend to discard many discriminative Gabor features while retaining several redundant ones during dimensionality reduction). However, it is more time consuming than either of these methods. To deal with this drawback, a hierarchical version of their approach is introduced. In this version, the face images are represented by an image pyramid (Figure 5.3(g)). Suppose the size of the face images is 2^Lx2^L. Then the pyramid can have a maximum of L+1 levels where the images at level i are of size 2^{i-1}x2^{i-1}, i=1, 2, … , L+1. The face recognition process starts at the topmost layer (with the smallest size images) and proceeds downward to the bottommost layer (with the largest size images). At each layer, the least similar images are discarded reducing the size of the gallery that the probe is compared against in the next layer. Hence, the computation time for the next layer will be comparable to the previous one, even though the classifier has become more complex. Thus, the hierarchical method should achieve both good accuracy and speed. The system was tested using images from the FERET database [32] which contain variations in expression, lighting and age. The non-hierarchical version of their approach with N=16 and 64 achieved significantly better results than GFC [357] as well as two other face recognition techniques, i.e., LBP [358] and LGBPHS [359] (which have reported the best results on the FERET database so far). The hierarchical version using 3 layers, where *N*=1, 4, 16 respectively, had significantly lower time cost than its non-hierarchical counterpart while achieving the same accuracy. Also, it had much higher accuracy than but took slightly more time than GFC [357].

5.3.1.1. Summary

The components used by the techniques discussed above are shown in Figure 5.3. A summary of these approaches in terms of the images databases they utilized, the algorithms used



(a) (b) (c) (d)

(e) (f) (g)

Figure 5.3. Facial components used by the following methods: (a) Price and Gee [23] (Reprinted with permission from Elsevier), (b) Toygar and Acan [24] (Reprinted with permission from Elsevier), (c) Yan and Osadciw [25] (©2005 IEEE), (d) Kumar et al. [26] (With kind permission of Springer Science and Business Media), (e) Zhang et al. [27] (With kind permission of Springer Science and Business Media), (f) Gottumukkal and Asari [28] (Reprinted with permission from Elsevier), (g) Su et al. [29] (©2006 IEEE).

for comparison, the factors that were tested and the rules used for combining the similarity scores is provided in Table 5.1.

Table 5.1. Summary of representative component-based approaches modeled on [1]

Approach	Face databases used for testing	Algorithms compared against	Factors tested	Combination rule(s)
Price and Gee [23]	CVL [32] and Yale [344]	PCA, PCA+LDA, D-LDA, DW-LDA	Expression, illumination, decoration	Sum rule
Toygar and Acan [24]	FERET [32]	Holistic PCA, LDA and ICA	Short aging, head rotation, illumination, scale	Sum rule (though other rules were also tested)
Yan and Osadciw [25]	Frontal images from ORL [342], AR [343], Yale [344], UMIST [349], FERET [32]	Holistic PCA	Expression, illumination, head leaning	Score level fusion (weighted sum), decision level fusion
Kumar et al. [26]	Yale [344]	Holistic PCA, modular PCA	Expression, illumination	Weighted sum rule
Zhang et al. [27]	Frontal images from FERET [32]	PCA+LDA	Expression	Concatenate the feature vectors obtained by applying PCA+LDA to each component and then apply LDA to the resulting vector
Gottumukkal and Asari [28]	UMIST [349] and Yale [32]	Holistic PCA, modular PCA	Expression, illumination, head pose angles	Sum rule
Su et al. [29]	FERET [32]	GFC, AGFC, LBP, LGBPHS	Expression, illumination, short aging	Sum rule

5.3.2. Methods not based on the modular eigenspaces approach

A brief discussion of some other techniques which are not based on Pentland et al.'s approach [1] is provided in this section. These techniques utilize methods other than PCA and LDA and have reportedly yielded encouraging results for the face recognition task.

Heisele et al. [30] use a SVM-based system to automatically detect and extract 14 facial components from face images, 10 of which are then used for classification (Figure 5.4). The components are then normalized in size and their gray values are combined into a single feature vector. For recognition, a linear SVM is associated with each person in the database where the SVM is trained to distinguish between all images of that person and all other images in the training set. The farther a probe image is from the decision surface found by the linear SVM associated with a certain person, the more similar it is considered to be to that person. Hence, the probe image is classified as the person for which that distance is the greatest. Two global systems are also implemented. One is the same as the component-based system described above except that it uses whole face images instead of facial components. The other first divides the images of each person into view-specific clusters. A linear SVM is then trained to distinguish between all images in one cluster and all images of other people in the training set. Tests were conducted on face data obtained from a video sequence. The training set consisted of 10000 facial images of 10 subjects (1400 were frontal views). Image resolution ranged from 80x80 to 130x130 pixels with rotations in azimuth up to about $\pm 40^{o}$. The test set contained 1544 images of the same subjects on a different day. Results indicate that the component-based system outperformed the global systems for recognition rates higher than 60%. However, for low recognition rates, it did slightly worse than the global schemes. This was attributed to the inability of the system to detect components correctly under large variations in pose. Also, the clustering method performed better than the other global technique. They conclude that "This shows that a combination of weak classifiers trained on properly chosen subsets of the data can outperform a single, more powerful classifier trained on the whole data." Others have also proposed component systems based on SVMs (e.g., [360]). However, a major limitation of such approaches, as pointed out in [30] is: “For a larger number of subjects, the choice of binary classifiers, like SVMs, might not be appropriate since the computational complexity for training and classification is linear with the number of classes.”.

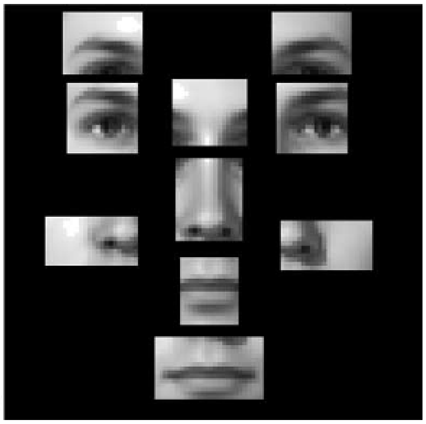


Figure 5.4. The 10 components used for classification by [30] (Reprinted with permission from Elsevier)

In Zana et al.’s [31] approach, the head is first detected in the image, the eye region is located and the eyes are found automatically. The image is normalized so that the eyes are registered at specific pixels. The images are cropped to 130x150 pixels, background information is removed and the resulting image is histogram equalized and normalized to zero mean and unitary standard deviation. Motivated by evidence that the human vision system extracts and utilizes global polar content of images [361], the face images are then transformed from the spatial to the polar domain by a Fourier-Bessel transform (FBT) [362]. FBT descriptors of the whole face as well as the 3 regions shown in Figure 5.5 are extracted. The pairwise Cartesian distances between all the FBT representations are then computed and each object is redefined by its distance to all other objects. Finally, the probe images are classified based on a pseudo FLD using a 2-class approach. The system was tested on the FERET database [32] using the gallery and 4 probe sets provided with it. The 4 algorithms used for comparison were Eigenfaces [6], EBGM [102], localized facial feature extraction followed by LDA [363] and a Bayesian generalization of the LDA method [364]. The systems' performance was compared under changes in expression, age, and illumination. For expression variations, the FBT approaches performed the same as the previous best and 2nd-best approaches. For age changes, they performed better than all previous methods. For illumination changes, they performed as well as or better than the second-best previous approach. Tests were also conducted to check the effects of occluding parts of the images. The component-based FBT method's performance degraded for the illumination dataset but remained about the same for the age and expression ones. However, the global FBT technique's performance degraded for all 3 sets. Others have also explored the viability of a component-based approach for dealing with occlusion [365, 366]. However, Zana et al. argue that their system is simpler in that it does not require detection of occluded regions, special training strategies or combination rules.

Figure 5.5. Sample of a normalized image and the regions used for local analysis [31] (©2005 IEEE)

A different technique which utilizes face segments is proposed by Nefian et al. [367]. In their method, an embedded HMM is constructed which consists of a set of super-states – each super state is itself a one-dimensional HMM [368]. Each individual is represented by an embedded HMM and facial regions are considered as super states in an embedded HMM as shown in Figure 5.6.

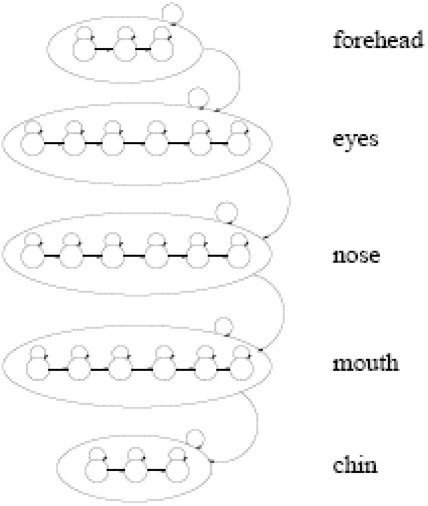


Figure 5.6. Embedded HMM for face recognition (©1999 IEEE)

Given an embedded HMM face model, the probability of the observation sequence is calculated using a doubly embedded Viterbi recognizer [369]. The model with the highest likelihood is then selected which reveals the identity of the probe face image. The system was tested on the ORL database [342] (40 people with 10 images per person displaying different expressions, hairstyles and eyewear). The system achieved a high recognition rate (98%) which was a 10% improvement over the one obtained by a one-dimensional HMM [370, 371]. It also performed better and was computationally less complex than the pseudo-2D HMM approach [215].

5.4. Conclusions

It would be hard to provide an objective comparison of all the component-based methods mentioned in this paper since the reported performance results evidently do heavily depend upon factors like the kind and number of images used, the image resolution, the amount of variation in properties such as age, orientation, expression, illumination and occlusion in the data, if and how the images were normalized, whether the components which were best for that specific approach were selected, whether the component extraction was manual or automatic (and thus, more prone to error), the values of the algorithm parameters, etc. As can be construed from Section 5.3, these key factors are different for each approach and thus, it is hard to arbitrarily judge if one method is better than another. However, the following general observations can be made based on the review in the preceding section:

- Most of these approaches do incorporate information about the whole face either by using the whole face as one of the components, or by selecting facial components that overlap. Failure to do so reportedly produces poorer results since information about the geometrical placement of facial features relative to each other is lost [1, 25]. The inclusion of holistic information for identification is also supported by evidence from psychological studies, which suggests that human beings recognize a face by first matching the whole face image and then further identifying it by some fine matching of the features [46].

- The eyes region has repeatedly been shown to be the best for discrimination relative to other prominent facial features (i.e., the nose and the mouth) [2, 23, 25].
- For approaches which explicitly extract certain regions from the face (rather than simply dividing it into equal-sized subimages), no methodical procedure has been described for choosing the components and no specific reason has been given for selecting a certain component set versus some other one. However, the components selected do generally contain features which are intuitively important for the human vision system (i.e., the eyes, nose and mouth), as pointed out by [1]: “Eye-movement studies indicate that these particular facial features represent important landmarks for fixation, especially in an attentive discrimination task [372].” More systematic approaches for selecting facial regions are being explored by others [373, 374].
- There appears to be a general consensus that component-based approaches perform better than holistic ones under variations in expression, illumination, decoration, head rotation and short aging. However, some results to the contrary have also been reported (e.g., Gottumukkal and Asari’s approach [28] did worse than holistic PCA under large pose variations, Toygar and Acan [24] found that their modular PCA method did worse than its holistic counterpart when several of the factors mentioned above were varied).

The results reported in the literature for various component-based approaches for face recognition indicate that, in general, such methods do perform significantly better than those which utilize the image of the whole face exclusively. This suggests that it would be worthwhile to invest some effort into further enhancing the performance of such techniques and/or incorporating such methods with other well-performing face recognition systems.

CHAPTER 6

A MULTI-RESOLUTION HIERARCHICAL APPROACH FOR FACE RECOGNITION[4]

---

[4] Rabia Jafri and Hamid R. Arabnia. 2007. The 2007 International Conference on Image Information and Knowledge Engineering (IKE'07), June 25-28, Las Vegas, USA. 231-239.
Reprinted here with permission of publisher.

Abstract

A hierarchical face recognition system which utilizes different resolution versions of the whole face, as well as various facial components, is proposed. Each level of the hierarchy uses different components and higher resolution images as compared to the previous layer and reduces the size of the gallery to be examined by the next layer. The idea here is to filter out the most similar faces at the first level using a minimum amount of information per image and then to feed those faces to subsequent levels each of which utilizes more information/image and returns a smaller (and hopefully, more accurate) set of the most similar images than the previous level. Experiments conducted to test the performance of the proposed system using frontal images from the FERET database and a 2D-PCA based face recognition method indicate that this technique performs better than its single-level counterpart which utilizes the highest resolution images.

Keywords: Face Recognition, Person Identification, Component-based, Resolution, Hierarchical.

6.1. Introduction

Multi-scale image pyramids have been extensively utilized in the field of computer vision for numerous applications, including image enhancement, object detection, pattern recognition, texture discrimination, and motion analysis, among others (please see [375, 376] for detailed reviews). In the area of automated face recognition, these representations have mainly been exploited for detecting the face and for speeding up the recognition process rather than for improving the accuracy of the results obtained. Our objective in this paper is to examine if using image pyramids along with the well-known concept of coarse-to-fine template matching [377] for the face recognition task would produce more accurate results as compared to those obtained by a system which uses only the highest resolution images.

We have also decided to utilize a component-based, rather than holistic, face recognition technique for our system. In contrast to most current automated face recognition approaches, which use the image of the whole face to perform classification, component-based techniques segment the face image into various parts, apply known face recognition methods to those facial regions and finally, combine the recognition results thus obtained to determine the identity of the whole face. These methods have been reported to produce significantly better recognition rates as compared to their holistic counterparts under variations in expression, illumination, decoration, head rotation and short aging [45].

An overview of the proposed system is as follows: Each face and facial component image is reduced to several lower resolutions. At the first level, recognition is performed using a certain subset of facial components and the lowest resolution versions of all images; the first x% closest matches are then selected as the new gallery. Each subsequent level, other than the last one, is similar to the first one except that it utilizes higher resolution images and finer components than the previous level. At the last level, the recognition results are returned. The idea here is to filter out the most similar faces at the first level using a minimum amount of information per image and then to feed those faces to subsequent levels each of which utilizes more information per image and returns a smaller (and hopefully more accurate) set of the most similar images than the previous level.

The rest of the paper is organized as follows: Section 6.2 discusses some related work done in the area of face recognition in terms of the use of image pyramids and facial components. Section 6.3 offers some insight into the motivation/reasoning behind the proposed approach. Section 6.4 describes the details of the system. Section 6.5 provides a description of the experiments conducted to evaluate this approach and an analysis of the results obtained. Section 6.6 contains some concluding remarks.

6.2. Related work

Since the proposed approach makes use of image pyramids and facial components for identifying faces, therefore, an outline of work related to these areas for the face recognition task will be presented in this section.

Face recognition systems typically include a face detection module which extracts the face from a scene and feeds it to the recognition module. Image pyramids have been widely used in these detection units to reduce the computational cost of locating the face in a scene as well as for dealing with variations in scale [378] [379]. Multi resolution image templates have also been exploited for speeding up the recognition process in object recognition systems [380-382]. However, the potential use of image pyramids to improve the accuracy, rather than the speed, of face recognition is an area that has not received much attention to date.

Component-based techniques for face recognition have become the focus of much research in the past few years. The earliest examples of such methods are the template matching strategy presented by Brunelli and Poggio [2] and the modular eigenspaces method introduced by Pentland et al. [1]. Most of the subsequent component-based approaches have been modeled on Pentland et al.'s method (for examples, please see [23] [26] [28]), though some other researchers have also utilized SVMs [30] [360] and embedded HMMs [367] for this purpose. Even though a few of these methods do divide the recognition process into stages (e.g., [27]

[353]), however, they do not use the intermediate stages to reduce the set of known individuals among which a match for the unknown person is to be found.

The only approach that we came across which is very similar to the one that we are proposing is that presented by Su et al. [29], since it uses both facial components as well as image pyramids and also employs a multi-stage recognition process where the number of candidates to be examined are reduced at each stage. However, their approach differs from the one proposed in this paper in that the facial components are obtained simply by dividing the whole face image into equal-sized subimages rather than by the selection of more meaningful facial regions (such as the eyes, nose etc.) which intuitively contain more discriminatory information as compared to other parts of the face. Furthermore, instead of working with the pixel intensity values, their technique extracts the Gabor features in each subimage and classification is performed using those features. Also, their recognition method uses a different dimensionality reduction technique than the one employed by our system (i.e., Fisher's linear discriminant method versus two-dimensional PCA in our case).

6.3. Motivation for a multi-resolution hierarchical component-based approach for face recognition

The proposed approach has two main characteristics:1)It utilizes facial components, rather than using the whole face alone, 2)It employs multi-resolution versions of images along with the concept of coarse-to-fine matching [377], performing recognition in a hierarchical fashion. The following discussion offers some insight into the motivation behind incorporating these two features into a face recognition system.

From a computational standpoint, methods that make use of facial components appear to offer several advantages over those that treat the whole face as one unit. Even though such techniques utilize the same algorithms that are employed by holistic approaches, however, since the statistical complexity of a facial component is less than that of the whole face image, therefore, applying a linear encoding via statistical techniques such as PCA/LDA/ICA to a facial region, rather than the whole face, is likely to result in better classification [159]. Also, pre-processing of small regions is easier in general than that of the whole image [353]. Furthermore, variations in expression, illumination and pose acutely affect some facial regions while leaving other segments of the face unchanged. The unchanged components will closely match their corresponding components in an image taken under normal conditions. Even for the affected regions, changes due to slight variations in pose or lighting will be smaller compared to the changes in the whole image pattern [30]. The same argument can be made for partially occluded or decorated (e.g., bearded or bespectacled) faces. Therefore, component-based schemes should perform better than their holistic counterparts in such scenarios. Moreover, Kim et al [353] argue that though variations in lighting and orientation would cause non-linear changes in the image intensity function for the whole face, however, in the face components, these effects can be approximated by a linear function, which can be corrected for much more easily. The reasons outlined above do appear to provide ample justification for a component-based approach for face recognition.

It also appears logical that dividing the recognition process into several levels, where each level examines a smaller set of candidates and processes a greater amount of information for each candidate than the previous level, would be beneficial for improving the recognition accuracy of the system. The first level utilizes the coarsest facial components and the lowest resolution images to rank the candidates in order of their similarity to the face of the unknown person. Since these images contain the least amount of data, and consequently, the least amount of discriminatory information, it is reasonable to assume that this ranking would not be very accurate and there is a good chance that the correct match would not be at the top of the resulting list. However, it is very likely that it would not be among the bottommost candidates on the list, either. Thus, if we select a major part of the list starting from the top, and discard the bottommost candidates, this would serve two purposes: 1) The most dissimilar candidates would have been removed from further consideration, and will, therefore, not confound the recognition results at latter stages. 2) The correct match would probably be among the selected candidates and since the selected candidates are roughly similar to each other based on their coarse features, the next rational step for discriminating among them would be to utilize more intricate versions of these images to examine their finer details. At the next level, the same process is repeated using yet finer facial components and higher resolution versions of the images except that a smaller number of candidates is retained at this stage, since there is a greater probability of the correct match being nearer the top of the list due to more information about it being available at this level. The process is repeated at subsequent levels so that at the last level, the set of candidate images of the highest resolution to be examined has the following two characteristics: 1)It is much smaller than the original candidate set, 2) It contains the images which are the most similar to the unknown face image and has a high probability of containing the correct match. Logic dictates that comparing the unknown face to this set is more likely to yield the correct match than comparing it to the original and much larger set of candidates.

It should be noted that the proposed hierarchical approach resembles the way in which human beings recognize faces. Results from several psychological studies indicate that human beings recognize a face by first matching the whole face image and then further identifying it by some fine matching of the features [46, 47]. It has also been found that the hair, face outline, eyes and mouth are important for perceiving and remembering faces [383] and that internal features (such as the eyes, nose and mouth) of familiar faces are more useful for recognition than external features (such as hair, jaw line) at high resolution but that the two feature sets reverse in importance as the resolution decreases [384]. The typical feature hierarchy reported in various studies is eyes, followed by the mouth and then the nose [385-387]. Furthermore, it has been reported that the upper part of the face is more beneficial for recognition than the lower part [383]. Thus, the proposed system emulates the human vision system in the following aspects: a) Using coarse/holistic features in the initial stages and finer internal features (such as the eyes, nose etc.) at the latter stages; b) Performing classification based on coarse holistic features from low resolution images and finer internal features from high resolution ones.

6.4. Details of the proposed approach

Let n be the number of resolutions and, consequently, the number of levels in the system. Let r_0 be the lowest resolution being used and let r_{i-1} be the image resolution being utilized at level i, where i=1, 2, … n, such that $r_{i-1} < r_i$. Let c_i denote the set of components being used at level i. Let m be the number of images in the original gallery and let x_i % of m gallery images be retained at level i.

A simple illustration of the proposed system is shown in Figure 6.1.

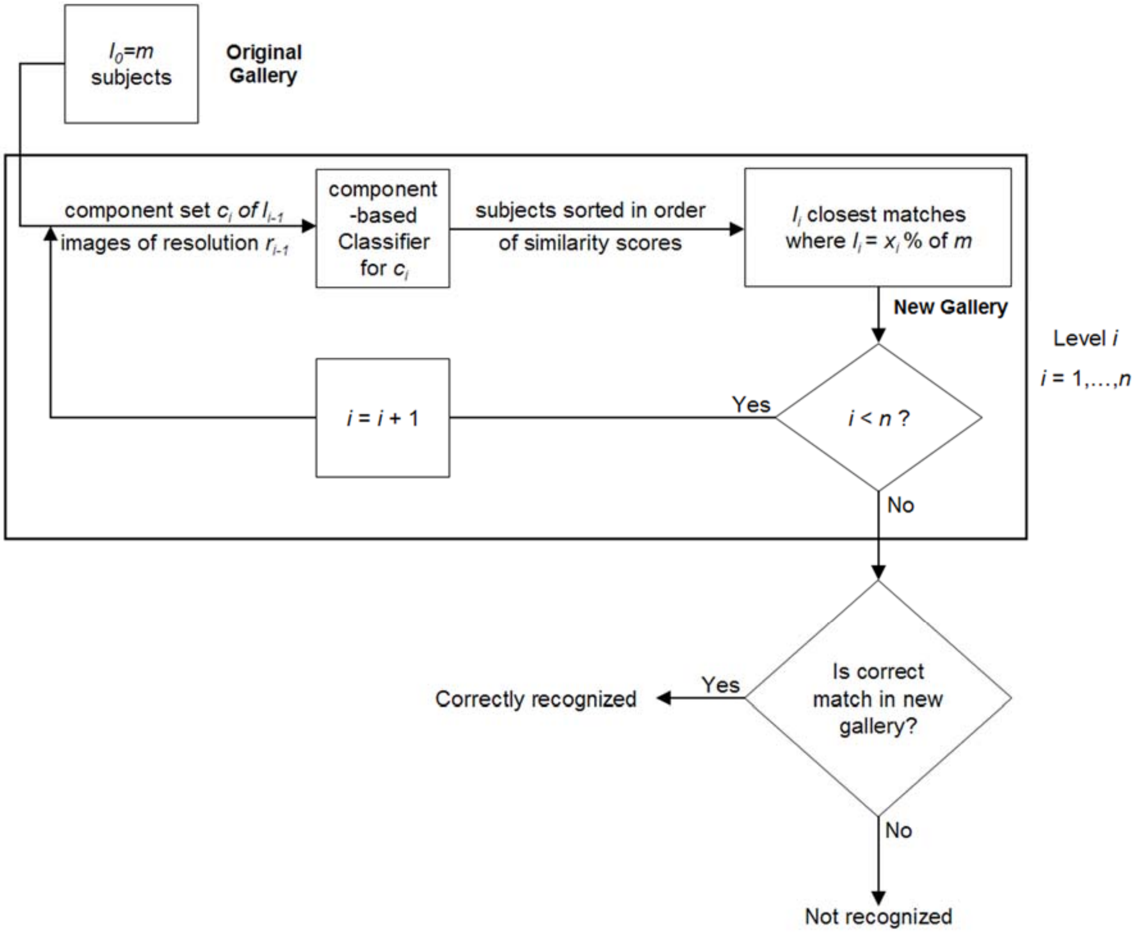


Figure 6.1. The framework for the proposed approach

6.4.1. Facial components

Taking our cue from psychology studies [383] [386], as well as previous results reported in the literature for automated component-based face recognition, we decided to utilize the following facial components for our system: the eyes region (note that the eyebrows are included in this area), the nose region, the upper part of the face containing both the eyes and the nose, and the whole face. The mouth was not selected as a separate component because this is a part of the face which changes the most due to variations in expression and it has been previously shown that its inclusion degrades the performance of automated systems for recognizing faces [1]. Figure 6.2 shows an example of facial components extracted from a face image.

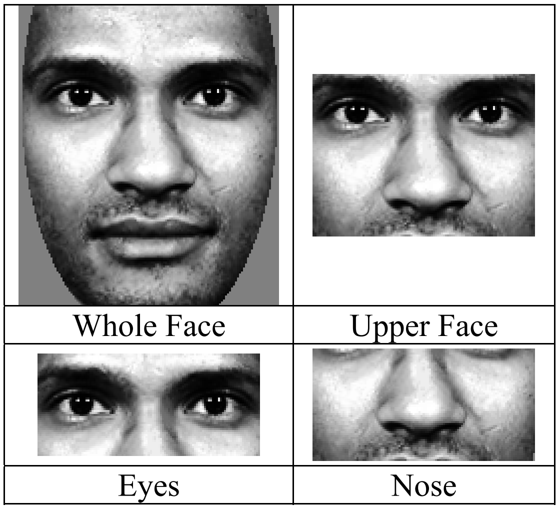

Whole Face
Upper Face
Eyes
Nose

Figure 6.2. Facial components extracted from a face image. (Face image taken from the FERET database [32])

6.4.2. Resolutions and facial components used at each level of the hierarchy

The original resolution of the face images was 256x384 pixels. The resolution of all the images was reduced to 128x 192, 64x96, and 32x48 respectively.

The resolutions and facial components utilized at each level of the hierarchy are shown in Table 6.1. Note that holistic features were used at the lowest resolutions while finer features were utilized at the higher ones. Since the eyes region has repeatedly been shown to be the best for discrimination relative to other prominent facial features (i.e., the nose and the mouth) [2, 23, 25], therefore, we decided to utilize this feature for classification at the last level.

Table 6.1. Resolutions and facial components utilized at each level of the hierarchy

Level	Resolution	Facial Components
1	32x48	Whole face
2	64x96	Whole face
3	128x192	Eyes + nose + upper face
4	256x384	Eyes

6.4.3. Face recognition method

The two-dimensional PCA-based method [341], a variant of the widely used PCA technique [6] for face recognition, was chosen to be the algorithm employed for classification. 2D-PCA has been shown to be as accurate as the one-dimensional PCA method, while computationally it is significantly faster [341, 388] (something which we verified through our own experiments, though the results are not depicted here).

6.5. Experiments

The following terminology introduced in [305] will be used in the rest of the paper: A *training set* is a collection of images used to construct the subspace; a *gallery* is a set of images of known individuals; a *probe set* consists of face images that need to be identified.

6.5.1. Face data and code

Experiments to test the above system were conducted on frontal face images from the FERET database [310]. The images of subjects wearing glasses were manually selected and discarded since it has been shown that glasses make subjects more recognizable [311]. A training set was then constructed from the resulting data set by selecting subjects with exactly three images. 203 such subjects (and hence, 609 images) were found. Out of the remaining images, 2289 images of 812 subjects were declared as potential candidates for inclusion in probe and gallery sets. Gallery image sets of sizes 100, 200, 300, 400, 500, 600, 700 and 800 were randomly selected from the available images such that no two images in one set were of the same subject. The corresponding probe sets of the same sizes were constructed by selecting one image of each subject in the gallery set such that that image was different from the one included in the gallery set.

The code for the chosen face recognition algorithms was acquired from the CSU face identification evaluation system [312]. The implementation details may be found in the user's guide for this system [313]. The code for the PCA algorithm was modified according to [341] to use 2D-PCA instead of traditional PCA for feature extraction.

6.5.2. Preprocessing the data, training and similarity score calculation

For images of each resolution, the following steps were carried out: The face images were normalized using the code provided in the CSU face identification evaluation system [312]. The centers of the eyes and nose were manually detected and the facial components described in section 6.4.1 were automatically extracted from the normalized images using these coordinates.

Training was then conducted for each set of components as described in [341]. 40% of the eigenvectors were retained in each case.

The gallery and probe image components were then projected onto their respective subspaces and the distances/similarity scores among them were calculated as explained in [341].

6.5.3. Recognition criteria

A probe image is considered to be correctly recognized if a correct match for it is found within the first 1% closest matches returned by the algorithm.

6.5.4. Results

6.5.4.1. Results for traditional component-based system using highest resolution images

The sum rule [315] was used to combine the normalized similarity scores of the highest resolution versions of the eyes, whole face, upper face and nose to obtain the final similarity score for a certain subject. The resulting recognition rates for various gallery sizes are shown in Table 6.2. The corresponding recognition rates obtained by using the whole face alone are also displayed here. The results confirm previous findings that the component-based approach performs better than the one which utilizes the whole face exclusively.

Table 6.2. Recognition rates for highest resolution using all the components versus the whole face alone

Gallery Size	Component-based	Whole Face
100	82.00%	76.00%
200	84.00%	83.00%
300	87.33%	84.00%
400	88.75%	84.25%
500	87.00%	84.00%
600	86.00%	81.83%
700	88.57%	81.71%
800	89.88%	85.88%

6.5.4.2. Results for multi-resolution component-based system

The multi-resolution system was tested using various percentages for the first 3 levels. 1% of the closest matches were retained at the last level to enable us to compare the recognition rates with those obtained using the highest resolution images alone. The results are depicted in Table 6.3.

Table 6.3. Recognition rates for multi-resolution and traditional component-based systems and the eyes classifier

Gallery Size	Multi-Resolution (Whole Face, Whole Face, Upper Face+Nose+Eyes, Eyes)						Traditional (highest resolution)	Eyes (highest resolution)
	(60%, 25%, 10%, 1%)	(50%, 25%, 20%, 1%)	(40%, 20%, 15%, 1%)	(30%, 15%, 10%, 1%)	(20%, 10%, 8%, 1%)	(10%, 8%, 5%, 1%)		
100	86.00%	86.00%	84.00%	84.00%	84.00%	84.00%	82.00%	85.00%
200	85.00%	84.00%	84.00%	84.00%	85.50%	86.00%	84.00%	83.00%
300	88.33%	87.67%	88.00%	88.33%	88.00%	88.33%	87.33%	87.00%
400	89.25%	89.00%	89.50%	89.25%	89.50%	88.50%	88.75%	88.00%
500	87.80%	87.80%	88.00%	88.00%	86.40%	86.80%	87.00%	86.80%
600	88.00%	87.67%	87.83%	87.67%	87.50%	87.00%	86.00%	86.67%
700	87.43%	86.86%	86.86%	87.00%	87.43%	87.71%	88.57%	86.57%
800	88.50%	88.13%	87.88%	88.00%	87.88%	88.50%	89.88%	87.63%

The following observations can be made based on these results:

- The multi-resolution system performs better in general than its traditional counterpart for all gallery sizes except for the two largest ones. The improvement in performance ranges from 0.25% to 4%.
- The multi-resolution system also yields better results in general than the eyes classifier (using the highest resolution images) alone. This shows that narrowing down the gallery at the lower levels does benefit the eyes classifier at the highest level to some extent.
- Decreasing the percentages of gallery images retained at the different stages alters the recognition rates somewhat but there is no consistent pattern and the variations in the recognition rates are 2 percentage points at the most.

We also tried using subsets of components different from the ones utilized above at the various stages. Some examples of the combinations used are given in Table 6.4. However, the results were almost the same or slightly worse than the combination used above, so there does not appear to be any significant advantage to using a different combination of the same components.

Table 6.4. Combinations of component subsets at various levels

Combination	Level			
	1	2	3	4
1	Whole face	Whole face	Upper face	Eyes
2	Whole face	Whole face	Upper face +eyes	Eyes
3	All components	Whole face	Upper face +eyes +nose	Eyes
4	All components	Whole face	Upper face	Eyes
5	All components	Upper face	Upper face +eyes +nose	Eyes

6.5.4.3. Results for variant of multi-resolution component-based system

The similarity score for each stage of the proposed system is based solely on the classifier in that stage and does not reflect information about the degree of similarity between the probe and gallery images obtained from the classifiers in the previous stages. To overcome this shortcoming, we decided to compute the similarity score at level *i* as follows:

$$similarity\ score\ (level\ i) = \sum_{j=1}^{i} normalized\ similarity\ score(level\ j) \tag{6.1}$$

All the tests in the previous section were conducted again using this new similarity score definition. Results are displayed in Table 6.5.

Table 6.5. Recognition rates for multi-resolution and traditional component-based systems and the eyes classifier

Gallery Size	Multi-Resolution (Whole Face, Whole Face, Upper Face+Nose+Eyes, Eyes)						Traditional (highest resolution)	Eyes (highest resolution)
	(60%, 25%, 10%, 1%)	(50%, 25%, 20%, 1%)	(40%, 20%, 15%, 1%)	(30%, 15%, 10%, 1%)	(20%, 10%, 8%, 1%)	(10%, 8%, 5%, 1%)		
100	85.00%	85.00%	85.00%	84.00%	84.00%	84.00%	82.00%	85.00%
200	88.00%	88.00%	88.00%	87.50%	87.00%	87.00%	84.00%	83.00%
300	89.00%	89.00%	89.00%	89.00%	89.00%	89.00%	87.33%	87.00%
400	90.00%	90.00%	90.00%	90.00%	89.75%	89.75%	88.75%	88.00%
500	88.60%	88.60%	88.40%	88.40%	88.20%	88.20%	87.00%	86.80%
600	87.67%	87.67%	87.67%	87.67%	87.50%	87.33%	86.00%	86.67%
700	88.86%	88.86%	88.71%	88.71%	88.43%	88.43%	88.57%	86.57%
800	90.50%	90.50%	90.50%	90.50%	90.50%	90.38%	89.88%	87.63%

The following observations can be made based on these results:

- Redefining the similarity score causes the recognition rates to be better than before for all but the smallest gallery size. This shows that taking classification information from previous stages into account does benefit the hierarchical system.
- The multi-resolution system now performs better than its traditional counterpart in all cases including the two largest gallery sizes. The improvement in performance ranges from 0.14% to 4%.

6.6. Conclusions

A multi-resolution component-based system was presented in this paper. The proposed system outperformed its traditional component-based counterpart for smaller gallery sizes. Redefining the similarity scores to take classification information from previous levels into account caused the system to yield better results than its traditional counterpart for larger gallery sizes, too. In our evaluation of the system, we used the same algorithm for images of all resolutions. Since some preliminary experiments that we previously conducted (please see [40] for details) have shown that certain subspace-based face recognition techniques perform better than others at higher resolutions and vice versa, hence, for future research, it would be interesting to investigate if using different algorithms for different resolutions in the proposed multi-resolution approach would further enhance the recognition performance. Furthermore, the highest resolution that we utilized was not very large (i.e., 256x384) and only a few major facial regions were used in our experiments. It would be worthwhile to examine how using yet higher resolution images and more facial components would affect the recognition accuracy of the proposed technique.

CHAPTER 7

A MULTI-LEVEL COMPONENT-BASED APPROACH FOR FACE RECOGNITION

7.1. Introduction

Face recognition is a challenging problem in the field of computer vision which has been the subject of extensive research in the past couple of decades because of its several applications in domains like security, surveillance, ubiquitous computing and video indexing [50, 51, 55, 57]. At present, most automated face recognition systems utilize the image of the whole face to perform classification [352]. However, in recent years, several techniques have been proposed which segment the face image into various parts, apply known face recognition methods to those facial regions and finally, combine the recognition results thus obtained to determine the identity of the whole face. These methods, generally referred to as component-based approaches, have been reported to produce significantly better recognition rates as compared to their holistic counterparts under variations in expression, illumination, decoration, head rotation and short aging [45].

A variation of a typical component-based approach is proposed in this paper whereas instead of examining the whole face and the facial components together at the same time in order to make a decision, the recognition process is divided into several levels as follows: At the first level, recognition is performed using the facial component which is the most pertinent to recognition and the entire database of known individuals is compared to the unknown person; the first *x*% matches are then selected as the new database of known subjects. At the second level, recognition is performed using the next most discriminating facial component and the database now is the one constructed at the previous level. Subsequent levels are similar to the second one except that at each level, the facial components considered are less pertinent to recognition than the previous level. The idea here is to use a relatively simple classifier to filter out the most similar faces at the first level and then to feed those faces to classifiers which look at the finer details at subsequent levels. Moreover, this appears to be in line with the way that human beings recognize faces [46, 47].

The rest of this chapter is organized as follows: Section 7.2 discusses some related work done in the area of facial component-based face recognition. Section 7.3 offers some insight into the motivation/reasoning behind the proposed approach. Section 7.4 describes the details of the proposed system. Section 7.5 provides a description of the experiments conducted to evaluate the proposed approach and an analysis of the results obtained. Section 7.6 contains some concluding remarks.

7.2. Related work

Several component-based techniques have been proposed in recent years. The earliest examples of such methods are the template matching strategy presented by Brunelli and Poggio [2] and the modular eigenspaces method introduced by Pentland et al. [1]. Most of the subsequent component-based approaches have been modeled on Pentland et al.'s method (for examples, please see [23] [26] [28]), though some other researchers have also utilized SVMs [30] [360] and embedded HMMs [367] for this purpose. Even though a few of these methods do divide the recognition process into stages (e.g., [27] [353]), however, they do not use the intermediate stages to reduce the set of known individuals among which a match for the unknown person is to be found. The only component-based strategy that we came across which does that is the one introduced by Su et al. [29]. However, their approach differs from the one proposed in this paper in that it uses different resolution versions of the same images at different stages and the facial components are obtained simply by dividing the whole face image into equal-sized subimages rather than by the selection of more meaningful facial regions (such as the eyes, nose etc.) which intuitively contain more discriminatory information as compared to other parts of the face. Given the lack of precedent for the unique strategy proposed in this paper, it appears worthwhile to investigate the feasibility of implementing it and to conduct some tests to determine if it works well in practice.

7.3. Motivation for a multi-level component-based approach for face recognition

The proposed approach has two main characteristics: 1) It utilizes facial components, rather than using the whole face alone, 2) The recognition process is divided into several levels, each level making use of a different subset of facial components. The following discussion offers some insight into the motivation behind incorporating these two features into a face recognition system both from a computational perspective as well as from a psychological/neurophysiological point of view.

From a computational standpoint, methods that make use of facial components appear to offer several advantages over those that treat the whole face as one unit. Even though such techniques utilize the same algorithms that are employed by holistic approaches, however, since the statistical complexity of a facial component is less than that of the whole face image, therefore, applying a linear encoding via statistical techniques such as PCA/LDA/ICA to a facial region, rather than the whole face, is likely to result in better classification [159]. Also, pre-processing of small regions is easier in general than that of the whole image region [353]. Furthermore, variations in expression, illumination and pose acutely affect some facial regions while leaving other segments of the face unchanged. The unchanged components will closely match their corresponding components in an image taken under normal conditions. Even for the affected regions, changes due to slight variations in pose or lighting will be smaller compared to the changes in the whole image pattern [30]. The same argument can be made for partially occluded or decorated (e.g., bearded or bespectacled) faces. Therefore, component-based schemes should perform better than their holistic counterparts in such scenarios. Moreover, Kim et al [353] argue that though variations in lighting and orientation would cause non-linear changes in the image intensity function for the whole face, however, in the face components, these effects can be approximated by a linear function, which can be corrected for much more easily. The reasons outlined above do appear to provide ample justification for a component-based approach for face recognition.

It also appears logical that dividing the recognition process into several levels should be beneficial for improving the recognition accuracy of the system. If an algorithm ranks the known individuals in order of similarity to an unknown person based on their most discriminating features, the assumption is that the individuals at the top of that ranking are more similar to each other as compared to those at the bottom of that list. Therefore, the subjects at the bottom of the ranking can be discarded as potential matches at that point. Since the remaining candidates already resemble each other strongly in terms of their most distinctive features, the logical course of action would be to compare finer details of their next most distinctive features and to rank them again based on that comparison. It is reasonable to assume that the set of individuals retained by extracting a percentage of the top candidates in the resulting ranking would be even more similar to each other in terms of their most distinctive features and looking at yet finer details of their less distinctive facial regions to discriminate among them would be the next rational step. This idea is related to the well known concept of coarse to fine template matching [377]. Of course, the success of the above scheme will depend heavily on the following two factors: 1) The hierarchy of facial components in terms of their importance for the recognition task is correctly determined, 2) The percentage of candidates is chosen at each stage in such a way so as to ensure that the correct match is not prematurely discarded.

The fact that several of the most widely used face recognition systems have been biologically inspired [107] lends credence to the notion of taking into account the recognition process of the human visual system while designing an automated system for this task. Exploring this avenue reveals strong support from several studies in cognitive psychology and neurophysiology for the use of facial components for face recognition by human beings as well as for there being a hierarchy in terms of the importance of various features for the recognition task. Results from several psychological studies indicate that human beings recognize a face by first matching the whole face image and then further identifying it by some fine matching of the features [46, 47]. It has also been found that the hair, face outline, eyes and mouth are important for perceiving and remembering faces [383] and that internal features (such as the eyes, nose and mouth) of familiar faces are more useful for recognition than external features (such as hair, jawline) at high resolution [384]. The typical feature hierarchy reported in various studies is eyes, followed by the mouth and then the nose [385-387]. Furthermore, it has been reported that the upper part of the face is more beneficial for recognition than the lower part [383].

7.4. Details of the proposed approach

Let $c_1, c_2, \dots, c_n$ be the facial components being used, arranged in order of importance for the face recognition task, with c_1 being the most important and c_n the least. Let m be the number of images in the original gallery and let x_i % of m gallery images be retained at level i, where $i=1, \dots, n$.

A simple illustration of the proposed system is shown in Figure 7.1.

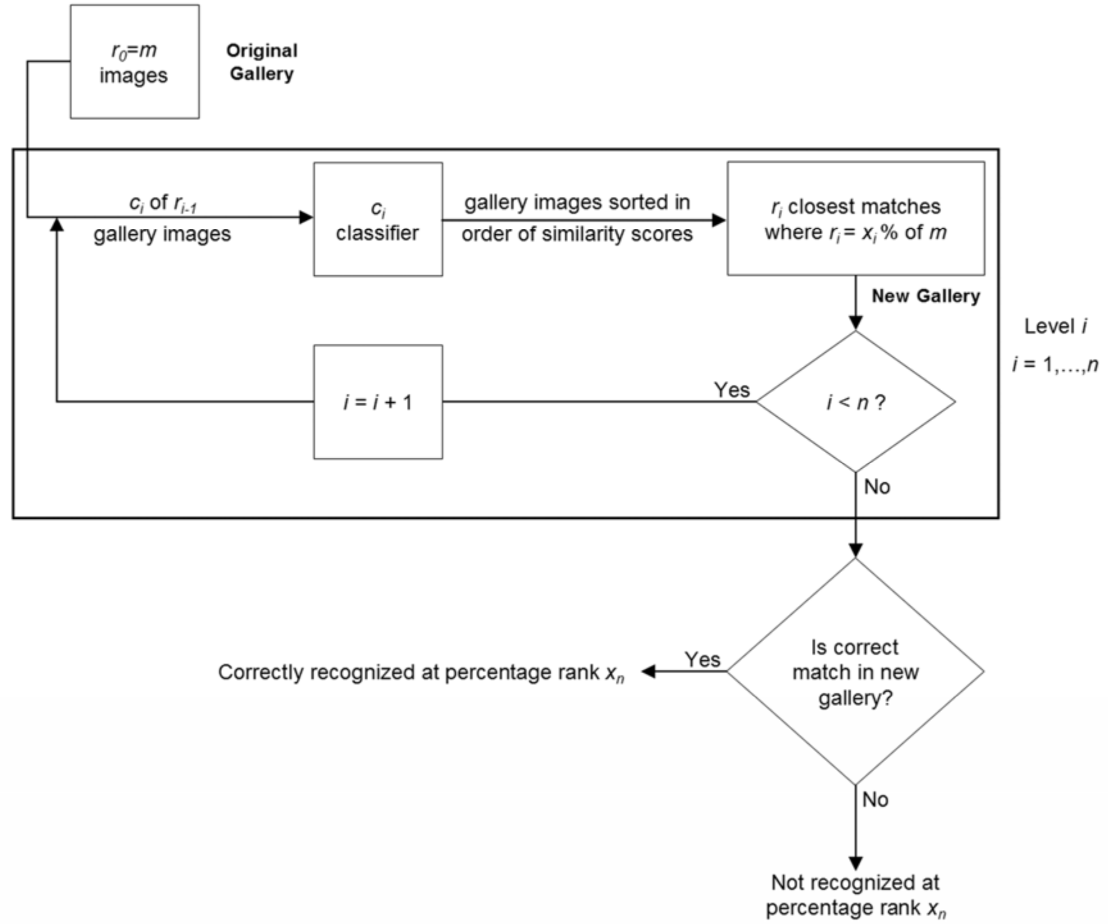


Figure 7.1. The framework for the proposed approach

7.4.1. Facial components

Taking our cue from psychology studies, as well as previous results reported in the literature for automated component-based face recognition, we decided to utilize the following facial components for our system: the eyes region (note that the eyebrows are included in this area), the nose region, the upper part of the face containing both the eyes and the nose, and the whole face. The mouth was not selected as a separate component because this is a part of the face which changes the most due to variations in expression and it has been previously shown that its inclusion degrades the performance of automated systems for recognizing faces [1]. Figure 7.2 shows an example of facial components extracted from a face image.

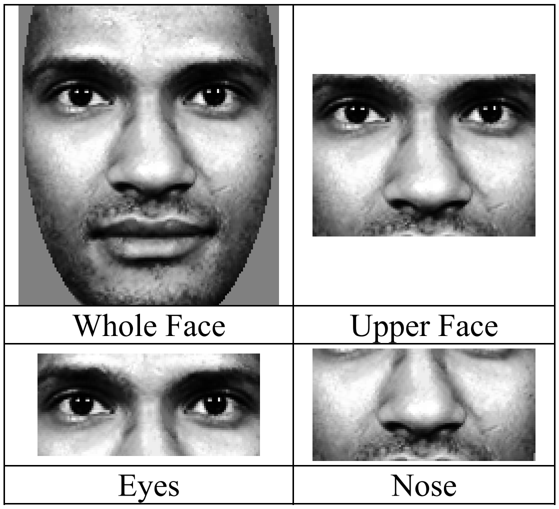


Figure 7.2. Facial components extracted from a face image. (Face image taken from the FERET database [32])

7h.4.2. Hierarchy of facial components

To determine the hierarchy of the facial components in terms of their importance for face recognition, we decided to adopt the following methodology: The recognition rates obtained by each individual component would be recorded. The components would then be ranked in descending order of their recognition rates.

7.4.3. Face recognition method

The two-dimensional PCA-based method [341], a variant of the widely used PCA technique [6] for face recognition, was chosen to be the algorithm employed for classification. 2D-PCA has been shown to be as accurate as the one-dimensional PCA method, while computationally it is significantly faster [341, 388] (something which we verified through our own experiments, though the results are not depicted here).

7.5. Experiments

The following terminology introduced in [305] will be used in the rest of the paper: A *training set* is a collection of images used to construct the subspace; a *gallery* is a set of images of known individuals; a *probe set* consists of face images that need to be identified.

7.5.1. Face data and code

Experiments to test the above system were conducted on frontal face images from the FERET database [310]. The images of subjects wearing glasses were manually selected and discarded since it has been shown that glasses make subjects more recognizable [311]. A training set was then constructed from the resulting data set by selecting subjects with exactly three images. 203 such subjects (and hence, 609 images) were found. Out of the remaining images, 2289 images of 812 subjects were declared as potential candidates for inclusion in probe and gallery sets. Gallery image sets of sizes 100, 200, 300, 400, 500, 600, 700 and 800 were randomly selected from the available images such that no two images in one set were of the same subject. The corresponding probe sets of the same sizes were constructed by selecting one image of each subject in the gallery set such that that image was different from the one included in the gallery set.

The code for the chosen face recognition algorithms was acquired from the CSU face identification evaluation system [312]. The implementation details may be found in the user's guide for this system [313]. The code for the PCA algorithm was modified according to [341] to use 2D-PCA instead of traditional PCA for feature extraction.

7.5.2. Preprocessing the data, training and similarity score calculation

The face images were normalized using the code provided in the CSU face identification evaluation system [312]. The centers of the eyes and nose were manually detected and the facial components described in section 7.4.1 were automatically extracted from the normalized images using these coordinates.

Training was then conducted for each set of components as described in [341]. The top 40% of the eigenvectors were retained in each case.

The gallery and probe image components were then projected onto their respective subspaces and the distances/similarity scores among them were calculated as explained in [341].

7.5.3. Recognition criteria

A probe image is considered to be correctly recognized at percentage rank x if a correct match for it is found within the first x% closest matches returned by the algorithm.

7.5.4. Determining the hierarchy of the components

The recognition rates obtained for each component for various gallery sizes at percentage rank 1 are shown in Table 7.1. The results indicate the following ranking of facial components in terms of their importance for recognizing faces: eyes, whole face, upper face and nose.

Table 7.1. Recognition rates for components (percentage rank=1)

Gallery Size	Eyes	Nose	Upper Face	Whole Face
100	85.00%	63.00%	77.00%	76.00%
200	83.00%	65.50%	80.00%	83.00%
300	87.00%	71.67%	82.67%	84.00%
400	88.00%	69.75%	83.25%	84.25%
500	86.80%	69.20%	80.60%	84.00%
600	86.67%	68.67%	80.33%	81.83%
700	86.57%	66.29%	82.43%	81.71%
800	87.63%	70.38%	83.50%	85.88%

7.5.5. Results

7.5.5.1. Results for traditional component-based system

The sum rule [315] was used to combine the normalized similarity scores of the eyes, whole face, upper face and nose to obtain the final similarity score for a certain subject. The

Table 7.2. Recognition rates using all the components versus those using the whole face alone

Gallery Size	Component-based	Whole Face
100	82.00%	76.00%
200	84.00%	83.00%
300	87.33%	84.00%
400	88.75%	84.25%
500	87.00%	84.00%
600	86.00%	81.83%
700	88.57%	81.71%
800	89.88%	85.88%

resulting recognition rates for various gallery sizes at percentage rank 1 are shown in Table 7.2. The corresponding recognition rates for the whole face alone are also displayed here. The results confirm previous findings that the component-based approach performs better than the one which utilizes the whole face exclusively.

7.5.5.2. Results for multi-level component-based system

The multi-level system was tested using various percentages for the first 3 levels. 1% of the closest matches were retained at the last level to enable us to compare the recognition rates with those of the traditional component-based approach as well as the nose classifier. The results are depicted in Table 7.3.

The following observations can be made based on these results:

1. The multi-level system performs worse than its traditional counterpart in all cases.
2. The recognition rates obtained are higher than those acquired by using the component at the last level (i.e., the nose) alone for classification. This shows that narrowing down the gallery using the eyes, whole face and upper face classifiers does assist the nose classifier to some extent.
3. The more the upper-level galleries are narrowed down, the better the recognition rates obtained at the last level are.

The above test will be referred to as test 1. The following additional tests were conducted to further analyze the system:

- Test 2: One obvious reason for the poor performance of the multi-level system could be that the correct matches are being prematurely discarded by the upper level classifiers, so that the classifier at the last level never gets the opportunity to select those. To verify if this is the case, the tests were run again with the same percentages at the first 3 levels. However, this time the nose classifier was not applied at the last level. The resulting recognition rates (shown in Table 7.4) are significantly higher than the percentage rank 1 recognition rates of the traditional component-based classifier. This confirms that the upper-face classifier in the 3rd stage is actually producing a gallery set which does contain most of the correct matches and that the nose classifier in the last stage is simply unable to find the best matches from that gallery.

Table 7.3. Recognition rates for 4-level and traditional component-based systems and the nose classifier at percentage rank = 1

Gallery Size	Multi-Level (Eyes, Whole Face, Upper Face, Nose)					Traditional (with 4 components)	Nose
	(50%, 25%, 20%, 1%)	(40%, 20%, 15%, 1%)	(30%, 15%, 10%, 1%)	(20%, 10%, 8%, 1%)	(10%, 8%, 5%, 1%)		
100	65.00%	65.00%	67.00%	68.00%	71.00%	82.00%	63.00%
200	66.50%	67.00%	67.50%	69.00%	71.50%	84.00%	65.50%
300	72.67%	73.33%	74.67%	75.33%	75.67%	87.33%	71.67%
400	70.75%	71.50%	71.50%	72.00%	73.25%	88.75%	69.75%
500	70.40%	71.00%	71.80%	72.80%	74.00%	87.00%	69.20%
600	69.33%	69.67%	70.17%	72.00%	73.00%	86.00%	68.67%
700	67.00%	67.71%	68.14%	69.71%	72.14%	88.57%	66.29%
800	71.50%	71.63%	72.13%	73.00%	75.00%	89.88%	70.38%

Table 7.4. Recognition rates for 3-level component-based system at various percentage ranks

Gallery Size	Multi-Level (Eyes, Whole Face, Upper Face)				
	(50%, 25%, 20%)	(40%, 20%, 15%)	(30%, 15%, 10%)	(20%, 10%, 8%)	(10%, 8%, 5%)
100	96.00%	96.00%	96.00%	96.00%	91.00%
200	98.00%	96.50%	95.00%	93.00%	91.50%
300	98.00%	97.00%	96.33%	96.00%	93.00%
400	98.50%	97.50%	95.75%	95.25%	92.25%
500	98.00%	97.00%	95.80%	95.40%	93.40%
600	98.50%	97.67%	95.67%	95.50%	92.83%
700	97.43%	96.57%	95.57%	95.57%	93.57%
800	98.13%	97.88%	96.88%	96.50%	94.25%

- Test 3: This is the same as test 1 except that the 4th level is discarded and 1% of the closest matches are retained at the 3rd level by the upper face classifier, enabling us to compare the recognition rates of the 3-level system with those of the single level one with 3 components at percentage rank 1. The results (shown in Table 7.5) indicate that the multi-level system still performs worse than the single level one. However, its performance is better than that of the 4-level one reinforcing the conclusion that the addition of the weak nose classifier at the 4th level just deteriorates the system's performance. The observations for tests 1 and 2 for the 4-level system also hold for the 3-level system with the following exceptions: Observation 2 for test 1 does not hold for the first three gallery sizes. Observation 3 for test 1 does not hold for the first 4 gallery sizes. However, it should be noted that the recognition rates of different percentage combinations for this system are very close to each other in all cases and that the differences in performance are less than a percentage point.
- Test 4: This the same as test 3 except that the 3rd level is discarded resulting in a 2-level system where the 2nd level classifier returns 1% of the closest matches (results are shown in Table 7.6). All observations for tests 1 and 2 for the 4-level system and test 3 for the

Table 7.5. Recognition rates for 3-level and traditional component-based systems and the upper face classifier at percentage rank = 1

Gallery Size	Multi-Level (Eyes, Whole Face, Upper Face)					Traditional (with 3 components)	Upper Face
	(50%, 25%, 1%)	(40%, 20%, 1%)	(30%, 15%, 1%)	(20%, 10%, 1%)	(10%, 8%, 1%)		
100	76.00%	76.00%	76.00%	76.00%	77.00%	83.00%	77.00%
200	80.00%	80.00%	79.50%	79.50%	80.00%	87.00%	80.00%
300	82.67%	82.67%	82.33%	82.33%	82.00%	89.33%	82.67%
400	83.75%	83.75%	83.75%	83.25%	83.00%	90.25%	83.25%
500	80.80%	81.00%	80.80%	80.80%	81.80%	88.00%	80.60%
600	80.83%	80.83%	81.50%	81.83%	81.83%	89.00%	80.33%
700	82.71%	82.71%	82.57%	83.14%	83.14%	89.57%	82.43%
800	83.75%	83.63%	83.63%	83.88%	83.88%	91.00%	83.50%

Table 7.6. Recognition rates for 2-level and traditional component-based systems and the whole face classifier at percentage rank = 1

Gallery Size	Multi-Level (Eyes, Whole Face)					Traditional (with 2 components)	Whole Face
	(50%, 1%)	(40%, 1%)	(30%, 1%)	(20%, 1%)	(10%, 1%)		
100	76.00%	76.00%	76.00%	77.00%	79.00%	86.00%	76.00%
200	84.00%	84.50%	84.50%	84.50%	85.50%	88.50%	83.00%
300	84.00%	84.00%	84.00%	84.00%	84.33%	89.33%	84.00%
400	84.25%	84.50%	84.50%	85.50%	86.25%	90.75%	84.25%
500	83.80%	83.80%	83.80%	84.00%	85.00%	88.20%	84.00%
600	81.83%	81.83%	82.33%	83.00%	84.00%	89.50%	81.83%
700	82.00%	82.00%	82.14%	82.43%	83.71%	90.29%	81.71%
800	85.88%	85.75%	85.88%	86.13%	86.63%	91.00%	85.88%

3-level system are true in this case, too.

Based on the above experiments, the following general conclusions can be drawn about the multi-level component-based scheme:

- It performs worse than its traditional counterpart in all cases.
- It yields better recognition rates than those obtained by using the component classifier at the last level alone. This shows that narrowing down the gallery using upper-level classifiers does benefit the lower level classifiers.
- The more the upper-level galleries are narrowed down, the better the recognition rates obtained at the last level are.
- A system with *x* levels performs better than that with *y* levels, where *x*<*y*. The apparent reason is that the classifiers at the lower levels are so weak that even given a small gallery containing most of the correct matches, they are still unable to find those matches and, therefore, degrade the accuracy of the system.

7.5.5.3. Results for variant of multi-level component-based system

The results of tests 2 and 3 in the previous section reveal a limitation of the multi-level approach in that the similarity score in each stage is based solely on the classifier in that stage and does not reflect information about the degree of similarity between the probe and gallery images obtained from the classifiers in the previous stages. To overcome this shortcoming, we decided to compute the similarity score at level *i* as follows:

$$\textit{similarity score}\ (\textit{level } i) = \sum_{j=1}^{i} \textit{normalized similarity score}(\textit{level } j) \qquad \textbf{(7.1)}$$

All the tests in the previous section were conducted again using this new similarity score definition. Results for 4, 3, and 2-level systems are displayed in Tables 7.7-7.9.

The following observations can be made based on the results:

- The observations for test 1 and test 2 for the 4-level system in the previous section are true in this case, too, even though narrowing down the upper-level galleries does not improve the recognition rates obtained at the last level as significantly as before.
- The observations for test 1 and test 2 for the 4-level system and test 3 for the 3-level system in the previous section are true for both the 3-level system and the 2-level system in this case, too, except that the recognition rates obtained are worse than or equal to the traditional component-based system. Again, narrowing down the upper-level galleries does not improve the recognition rates obtained at the last level as significantly as before.

The following conclusions can be drawn based on the above observations:

- Redefining the similarity score improves the performance of the multi-level system.
- The multi-level system, that uses the new definition of the similarity score (which takes the upper level classifiers' results into account), is not able to perform better than the traditional component-based system.

Table 7.7. Recognition rates for 4-level and traditional component-based systems and the nose classifier at percentage rank = 1

Gallery Size	Multi-LeveL (Eyes, Whole Face, Upper Face, Nose)					Traditional (with 4 components)	Nose
	(50%, 25%, 20%, 1%)	(40%, 20%, 15%, 1%)	(30%, 15%, 10%, 1%)	(20%, 10%, 8%, 1%)	(10%, 8%, 5%, 1%)		
100	79.00%	79.00%	79.00%	80.00%	80.00%	82.00%	63.00%
200	82.50%	82.50%	82.50%	83.00%	82.50%	84.00%	65.50%
300	85.67%	85.67%	85.67%	85.67%	85.33%	87.33%	71.67%
400	87.25%	87.25%	87.25%	86.75%	86.50%	88.75%	69.75%
500	85.00%	85.20%	85.20%	85.00%	85.20%	87.00%	69.20%
600	83.67%	83.67%	83.67%	83.83%	84.00%	86.00%	68.67%
700	86.14%	86.29%	86.29%	86.29%	86.14%	88.57%	66.29%
800	88.00%	88.00%	88.00%	87.88%	87.88%	89.88%	70.38%

Table 7.8. Recognition rates for 3-level and traditional component-based systems and the upper face classifier at percentage rank = 1

Gallery Size	Multi-Level (Eyes, Whole Face, Upper Face)					Traditional (with 3 components)	Upper Face
	(50%, 25%, 1%)	(40%, 20%, 1%)	(30%, 15%, 1%)	(20%, 10%, 1%)	(10%, 8%, 1%)		
100	83.00%	83.00%	83.00%	83.00%	83.00%	83.00%	77.00%
200	87.00%	87.00%	87.00%	86.50%	86.50%	87.00%	80.00%
300	89.33%	89.33%	89.33%	89.33%	89.00%	89.33%	82.67%
400	90.25%	90.25%	90.25%	89.75%	89.25%	90.25%	83.25%
500	88.00%	88.00%	88.00%	88.00%	88.20%	88.00%	80.60%
600	89.00%	89.00%	89.00%	89.00%	88.67%	89.00%	80.33%
700	89.57%	89.57%	89.57%	89.57%	89.29%	89.57%	82.43%
800	91.00%	91.00%	91.00%	90.88%	90.38%	91.00%	83.50%

Table 7.9. Recognition rates for 2-level and traditional component-based systems and the whole face classifier at percentage rank = 1

Gallery Size	Multi-Level (Eyes, Whole Face)					Traditional (with 2 components)	Whole Face
	(50%, 1%)	(40%, 1%)	(30%, 1%)	(20%, 1%)	(10%, 1%)		
100	86.00%	86.00%	86.00%	87.00%	86.00%	86.00%	76.00%
200	88.50%	88.50%	88.50%	88.50%	88.50%	88.50%	83.00%
300	89.33%	89.33%	89.33%	89.33%	89.33%	89.33%	84.00%
400	90.75%	90.75%	90.75%	90.50%	90.00%	90.75%	84.25%
500	88.20%	88.20%	88.20%	88.20%	88.00%	88.20%	84.00%
600	89.50%	89.50%	89.50%	89.50%	89.50%	89.50%	81.83%
700	90.29%	90.29%	90.29%	90.29%	89.86%	90.29%	81.71%
800	91.00%	91.00%	91.00%	90.88%	90.63%	91.00%	85.88%

- The addition of significantly weak classifiers (such as the nose classifier) at the lower levels can offset all the good done by the higher level classifiers, i.e., even if the higher level classifiers narrow down the gallery so that it contains almost all the correct matches along with a few incorrect ones, the lower level classifier would inject so much inaccuracy in the similarity scores so as to cause incorrect matches to be given precedence over the correct ones. This apparently is the key reason for the failure of this approach to perform any better than its traditional component-based counterpart.

- Narrowing down the gallery at the upper levels too much degrades the recognition accuracy, probably, because this causes some of the correct matches to be prematurely discarded.

7.6. Conclusions

A multi-level component-based system was presented in this chapter. The initial version of this system failed to produce better results than its traditional single-level counterpart. Though changing the similarity score definition so that it reflected the classification information from all previous stages resulted in the multi-level system performing better than before, however, it was still unable to yield better recognition rates than the traditional component-based one. The key reason for the failure of this system appears to lie in the inability of the classifiers at the lowest levels to find the correct match even when presented with a smaller set of candidates containing that match. In our evaluation of the system, only a few major facial regions were utilized. For future research, it would be interesting to investigate if employing a more systematic way of selecting a different set of facial components (see [373, 374] for examples) and using 2 or more components at each level, rather than just one, might boost the recognition accuracy.

CHAPTER 8

FUSION OF FACE AND GAIT FOR AUTOMATIC HUMAN RECOGNITION[5]

---

[5] Rabia Jafri and Hamid R. Arabnia. Proceedings of the 5th International Conference on Information Technology - New Generations (ITNG 2008; Data Mining Track), April 7-9, Las Vegas, USA; IEEE Computer Society. 167-173.
Reprinted here with permission of publisher.

Abstract

In recent years, several techniques have been proposed which integrate face, a physical biometric, with gait, a behavioral biometric, with the aim of investigating if such a combination will improve upon the performance of methods which exclusively employ only one of these biometrics. An overview of some of the well-known approaches in this area, along with a discussion of the advantages offered and the challenges faced by such systems, is provided in this paper and the potential of this technology for further research and application is explored.

8.1. Introduction

Face recognition is a challenging problem in the field of computer vision which has been the subject of active research for the past several decades because of its many applications in domains such as surveillance, covert security and context-aware environments. The face's appeal as a biometric stems from the several advantages that it offers in terms of being non-intrusive, non-invasive, cost-effective, easily accessible (i.e., face data can be conveniently acquired with a few inexpensive cameras) and relatively acceptable to the general public. However, employing the face for recognition also presents some difficulties since the appearance of the face can be altered by intrinsic factors such as age, expression, facial paraphernalia (facial hair, glasses, cosmetics, etc.), ethnicity, and gender as well as extrinsic ones such as illumination, pose, scale and imaging parameters (e.g., resolution, focus, imaging noise, etc.).

Gait is a behavioral biometric, whose utility for human recognition has only recently begun to be explored. Similar to the face, gait, too, is a visual cue which can be extracted from video and thus, it appears to offer the same advantages (outlined above) that are presented by the face. However, gait offers some additional benefits in that, unlike the face, it can be acquired from a sequence of low resolution images of a person taken from a distance where the subject's body occupies too few pixels for other biometric traits to be discerned. Moreover, it is more robust to slight variations in viewpoint as compared to other biometrics and cannot be easily disguised without attracting attention. Nevertheless, gait as a biometric has its limitations since, being a behavioral trait, it can be affected and altered by factors such as clothing, footwear, environmental conditions, emotions, fatigue, drunkenness, pregnancy, injury, disease, aging and load. In addition, gait also suffers from the usual problems associated with extracting visual cues from video such as imperfect foreground segmentation of the walking subject from the background scene, and poor imaging conditions.

In recent years, some approaches have emerged which integrate face, a physical biometric, with gait, a behavioral biometric, with the aim of investigating if such a combination will improve upon the performance of techniques which exclusively employ only one of these biometrics. Though the development of this multi-biometric combination can still be considered to be in its nascent stages, with relatively few papers having been published on this topic to date, nevertheless, the results reported so far are very encouraging and clearly demonstrate the potential of such an approach for discriminating among people. A survey of integrated face-gait recognition techniques has, therefore, been undertaken in this paper. The objective is to provide an overview of the state of the art in this area, highlighting the advantages offered by as well as the challenges faced by such systems, and to inspire and facilitate further exploration of this avenue of research.

The rest of this paper is organized as follows: Section 8.2 provides a brief overview of previous work in the areas of face, gait and multi-biometric recognition. Sections 8.3 offers some insight into the motivation for considering the integration of the face and gait biometrics for human recognition. Section 8.4 contains a review and analysis of some recently proposed methods for combining the face and gait. Section 8.5 includes some concluding remarks.

8.2. Related work

Face recognition has been the focus of extensive research for the past three decades (see [352] for a detailed survey). The approaches for this task can be broadly divided into two categories: 1) Feature-based methods [101, 102], which first process the input image to identify and extract distinctive facial features such as the eyes, mouth, nose, etc. as well as other fiducial marks and then compute the geometric relationships among those facial points, thus, reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed for matching faces using these measurements. 2) Appearance-based (or holistic) methods [6, 8], which attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. Though face recognition methods traditionally operate on static intensity images, in recent years, much effort has also been directed towards identifying faces from video [254] as well as from other modalities such as 3D [286] and infra-red [300].

Gait is a relatively new biometric characteristic. Methods to recognize gait from video sequences first emerged in the early 1990s mainly because it was around this time that computer memory and processing speed became sufficient to process large amounts of image data with reasonable efficiency [39]. Approaches that identify people from gait using video can broadly be partitioned into two groups [43]: 1) Model-based methods [42-44] use a model of either the person's shape or motion to derive static (i.e., absolute distances between certain body landmarks, such as the height, limb lengths, shoulder width, torso length, etc.) as well as dynamic gait features (e.g., swings of legs and arms, stride length, walking speed, and kinematics of joint angles) 2) Holistic techniques [389-391] represent the gait by the statistics of the spatiotemporal patterns generated by the walking person in an image sequence. It should be noted that some approaches have also been proposed to extract gait features from modalities other than video (e.g., Middleton et al. [392] utilize floor sensors to extract stride parameters as well as time on toe to time on heel ratios).

Recently, much effort has been expended on combining various biometrics in a bid to improve upon the recognition accuracy of classifiers that are based on a single biometric. Some biometric combinations which have been tested include face, fingerprint and hand geometry [393], face, fingerprint and speech [394], face and iris [395], face and ear [74], and face and speech [315]. The combination of face and gait being considered in this paper is just one more example of such a system.

8.3. Motivation for integrating face and gait

Since the performance of any classifier is more sensitive to some factors and relatively invariant to others, a recent trend has been to combine individual classifiers in order to integrate their complementary information and, therefore, create a system that is more robust than any individual classifier to variables that complicate the recognition task. Hong et al. [396] have shown that integrating multiple biometrics does indeed result in consistent performance improvement while Schiele et al. [397] have empirically demonstrated that as the number of classifiers combined increases, so does the recognition accuracy. The encouraging results reported in the literature for such systems [74, 315, 393-395] in conjunction with the conclusions reached by the studies mentioned above provide a strong basis for believing that a system constructed by combining various biometric characteristics is going to yield better recognition rates than the individual classifiers for those traits.

Furthermore, using multiple biometrics is a viable solution to real-world problems, such as non-universality of some biometric traits (e.g., some people's fingerprints cannot be reliably extracted because of the poor quality of the ridges), unavailability of data for a certain biometric (e.g., visual cues such as face, ear, etc. might be occluded in surveillance videos) and criminal activity (i.e., attempts to fool the single-biometric based system by duplicating the biometric trait or breaching the system).

In light of the above, some specific reasons for considering the integration of the face and gait biometrics, in particular, are as follows: the face is a short-range biometric, which can be used effectively for identification only when the subject is close enough to the camera for sufficient details of his facial features to be captured. Gait, on the other hand, is a medium to long-range biometric, which can be extracted reliably even from low-resolution imagery and is more invariant to slight changes in viewpoint. Using these two biometric traits together would arguably make the system more robust to variations in subject to camera distance. Also, both face and gait are visual cues; both can be extracted from the same modality, (i.e., image sequences of people) precluding the need for separate or specialized equipment. Furthermore, these biometrics make use of apparently independent personal characteristics: face recognition systems exploit the relatively detailed appearance of the facial surface, while gait recognition methods capture data from the coarse body shape as it changes over time. Consequently, some conditions that sharply degrade the performance of face recognition systems, such as large variations in illumination and facial expressions, affect gait to a much lesser extent or not at all. Similarly, some conditions that adversely affect the accuracy of gait recognition, such as clothing, footwear, and load, do not influence the performance of face recognition systems. Therefore, it is reasonable to believe that combining these complementary cues would improve the recognition accuracy.

8.4. Approaches integrating face and gait

Though the combination of the face with biometrics such as speech [315], ear [74], iris [395], etc., has been studied in some detail, however, the integration of the face with gait for human recognition is a relatively unexplored area. We first discuss various general schemes used to fuse biometrics before embarking on a review of combined face-gait systems.

Fusion of biometrics can be done at three levels [398]: 1) The feature level, where feature vectors of the individual biometrics are concatenated to form a new feature vector, which represents the subject's identity in a higher-dimensional and, presumably, more discriminating hyperspace. 2) The matching score level, where the similarity scores of the individual biometrics are first transformed (using transformations such as linear, exponential, logarithmic, etc.) to ensure that they map to a common range and then combined using some combination rule (e.g., SUM, PRODUCT, MAX, etc.) to obtain a new similarity score. 3) The decision level, where each individual classifier makes its own decision regarding the identity of the unknown person and some rule (e.g., majority voting, rank rule, etc.) is employed to make the final decision. Hierarchical systems, in which a weak classifier is used to prune off the most unlikely candidates and the remaining candidates are passed to a more powerful classifier which makes the final decision, may be assigned to this category.

Though most of the integrated face-gait systems proposed so far make use of hierarchical and matching-level fusion [399] [400], nevertheless, a couple of methods also utilize feature-level fusion [401] [34]. We now present an overview of some of the most well-known combined face-gait systems. A summary and comparison of these techniques is provided in Table 8.1.

Zhou et al. [402] utilize face profile and gait silhouette from single camera video sequences for recognition of people at a distance. First, a high resolution face profile is constructed for each subject from six adjacent low-resolution video frames using the super-resolution technique proposed by Irani and Peleg [403]. A curvature-based matching method [404] is then applied for face profile recognition. Next, the gait silhouette sequence extracted from the video sequence is normalized and the average of all the frames in the normalized sequence is used to construct a gait energy image (GEI). Gait recognition is performed by direct matching of the GEIs. Finally, the face profile and gait silhouette are combined using 3 methods: the SUM rule, the PRODUCT rule and a hierarchical scheme. For the first two fusion methods, the similarity scores from the two classifiers are first normalized using exponential transformation and are then combined using the relevant rule. The third scheme is a hierarchical one in which the top matches of the face profile classifier are passed on to the gait classifier which then proceeds to make the final decision. The system was tested using 14 people. The results indicate that the SUM rule performed the best, yielding 100% correct classification, while both the PRODUCT and hierarchical methods attained recognition rates of 92.9%.

Three variants of the above system are presented in later papers, which differ from it in the following ways: 1) They utilize the side face instead of the face profile. 2) Instead of using the enhanced side face image (ESFI) and GEI directly, they form feature vectors as follows: a) In [405], Principal components analysis (PCA) followed by multiple discriminant analysis (MDA) is applied to the ESFIs and GEIs to obtain the face and gait feature vectors; b) In [401], PCA+MDA is applied to the ESFIs and the GEIs and the resulting vectors are concatenated to form combined feature vectors; c) In [34], PCA is applied to the ESFIs and the GEIs; the resulting vectors are concatenated and MDA is applied to them to obtain the final combined feature vectors. In [405], classification is performed using individual face and gait classifiers whose results are combined using SUM and MAX rules, while a one-nearest neighbor classifier is utilized in [401] and [34]. These systems were tested and compared on a database of 45 people [34]. Two experiments were conducted: in the first, the gallery and probe sequences were taken the same day and there was no change in the subjects' clothing; in the second, for 10 of the subjects, the gallery and probe sequences were taken 1 month apart with different clothing. The recognition rates obtained are shown in Table 8.1. The results indicate that a)The fused techniques perform better than the individual face and gait classifiers in all cases, b)The feature-level fusion approaches [34, 401] perform better than the match-level fusion ones [405] in all cases, c) The performance of both feature-level fusion methods is the same in experiment 1 but [34] performs better than [401] in experiment 2.

Liu et al. [399] integrate frontal face and gait silhouette cues with the aim of investigating if doing so would improve recognition at a distance outdoors and over time, both of which are hard conditions. For gait recognition, a new algorithm is presented: Given a gait sequence, covering multiple gait cycles, the prominent stances (specifically those that have the largest variance across a population) are identified using a population Hidden Markov model (HMM). Eigenstance shape models are used to build averaged representations for the detected silhouettes for those stances. The similarity score between a gallery and a probe sequence is then computed by taking the sum of the Euclidean distances between the averaged representations for the selected stances. For face recognition, the Elastic Bunch Graph Matching [102] algorithm is employed. Scores from both classifiers are then normalized to a common range using Gaussian model-based z-normalization, which was also used in FRVT-2002 [86]. One decision level (based on negated sum of ranks) and three score level (based on the SUM rule, Bayesian decision rule and Confidence Weighted Score Sum) schemes are used to fuse the results of the two classifiers. The system was evaluated using a gallery of 70 individuals and two probe sets: one with 39 subjects taken the same day and the other with 21 individuals taken at least 3 months apart. For the same day probe set, the best recognition rate of 71% was obtained using the SUM rule while for the probe set taken months apart, the best recognition rate of 55% was obtained using the weighted score sum. Further experiments also showed that the multi-biometric combination (face+gait) was better not only than the individual biometrics but also combinations of the same biometric (i.e., face+face and gait+gait).

Kale et al. [400] observe that different viewing angles are desirable for face and gait recognition: a frontal view works best for the face while a side view is most advantageous for the gait. They, therefore, propose a method for view-invariant gait recognition using a single camera and proceed to integrate it with frontal face recognition. For their system, they utilize outdoor video sequences from the NIST database [406] in which subjects walk along an inverted Σ path. The top horizontal segment of the Σ where the subject walks at an exact side-angle to the camera is used for the gait gallery. The next segment in which the subject walks at an angle 33^{o} to the horizontal part is used as the gait probe, while the last part of the sequence where the person presents a frontal face view to the camera is utilized as the face probe. Static face images are employed for the face gallery. Their gait recognition strategy works as follows: The probe video sequences are converted to a side perspective by synthesizing side views of the person from arbitrary views using a transformation derived for this purpose. A template matching technique based on dynamic time warping [407] is then used for recognizing the gait by utilizing these synthetic sequences. Similarity scores are reported in terms of the cumulative binary correlation distances between the probe and gallery video sequences. For face recognition, a still-to-video sequential importance sampling based algorithm [408] is employed which yields similarity scores in terms of posterior probabilities. Since the gait recognition scores are in terms of distances while those of the face classifier are in terms of probabilities, an exponential transformation is applied to the gait scores and the transformed scores are then normalized to sum to unity in order to make them comparable with the face similarity scores. Two fusion scenarios are considered: hierarchical and holistic. In the hierarchical approach, the top matches from the gait classifier are passed to the face recognition algorithm which then makes the final classification decision. The main advantage offered by this strategy is the reduction in the total number of computations required. In the holistic method, the SUM or PRODUCT rule is used to combine the face and gait similarity scores. Tests were conducted using data for 30 people from the NIST database. The hierarchical scheme yielded better recognition rates as compared to those obtained using the face alone (97% vs. 93%), thus, demonstrating the effectiveness of gait as a filter. The holistic approach produced 100% recognition rates for both the SUM and the PRODUCT rules.

Shakhnarovich et al. [33] also address the issue of recognition from arbitrary views as well as that of recognition at a distance. To resolve the latter issue, they propose combining face and gait cues, and for the former one, they suggest computing an image-based 3-D visual hull (VH) from a set of monocular views and using that to render virtual views in canonical poses: frontal for the face and side-view for the gait. Image sequences from these virtual canonical views are then passed to the face and gait recognition units. Face recognition is done using the eigenfaces approach [6]. For gait recognition, the gait feature vector is constructed from a sequence of silhouettes by dividing each silhouette frame into regions and then taking smoothed versions of moment features in those regions. A nearest neighbor classifier is used to determine which person has walking dynamics closest to the query feature vector. Face and gait cues are combined by taking a weighted sum of their confidence vectors. Note that if no faces are detected in the synthetic frontal views, then recognition is based on gait alone. The system was tested using 12 subjects, with between 2 to 6 VH sequences per subject, and yielded higher recognition rates (91%) as compared to the individual face (80%) and gait (87%) classifiers. In a subsequent paper [409], this system was extended, with a more sophisticated probabilistic approach being used for combining the face and gait cues. Various fusion strategies (i.e., MEAN, MAX, MIN and MAJORITY rules) were empirically tested on a database of 26 subjects, where data for 11 of the subjects was collected on two separate days 3 months apart. All fusion methods were found to perform better than the individual face and gait classifiers with the best recognition rates being achieved by the PRODUCT rule.

All the methods mentioned so far employ holistic techniques for recognizing the gait. Lee et al. [410], however, opt to characterize the gait in a novel way by considering the time series of point coordinates generated by the motion of Moving Light Displays attached to various points of a walking person (i.e., hands, knees, feet, face and waist) when viewed from a frontal-normal perspective. Phase space analysis of that motion reveals chaotic behavior. The chaotic behavior is measured with Lyapunov exponents λ_1 calculated using the method proposed by Rosentein et al. [411]. The average of the exponents $\overline{\lambda_1}$ for all the body points of a person is used to characterize his gait. Similarity between two gaits is then measured by calculating the Jeffries-Matusita Distance (JMD) [412] between their $\overline{\lambda_1}$s. For face recognition, the gross location of the face is found using color and that region is then scanned for the exact face location using the concept of distance-from-face-space as described in [1]. Recognition is performed using the eigenfaces approach [6]. Gait and face cues are fused using a hierarchical scheme where the top matches of the gait recognition system are passed on to the face recognition unit which makes the final decision. The system was tested on 12 subjects. Actual results are reported for only three subjects: these show that the error rates of the fused system are significantly lower than those of the gait recognition system alone. However, the exact overall recognition rates have not been reported. More importantly, it has not been clarified if the fused system performs better than the individual face classifier, leading one to question if there is any merit to using the proposed scheme versus the face recognition system alone.

8.4.1. Summary

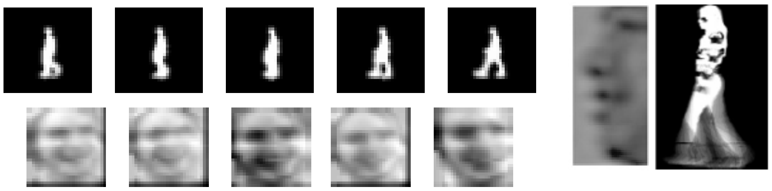
A summary of the various integrated face gait approaches discussed above in terms of the biometric features utilized, the data used for evaluating these strategies, the fusion methods employed, and the recognition rates obtained is provided in Table 8.1. Examples of the face and gait representations used by some of these techniques are given in Figure 8.1.

8.5. Conclusion

It would be hard to objectively compare the performance of the approaches discussed above since the recognition rates reported for these depend heavily upon factors such as the kind and the detail of biometric features utilized, how accurately those features were extracted, the number of test subjects and the amount of data available for each, the fusion technique, the parameter settings for the algorithms employed (e.g., for the eigenfaces method: the number of eigenvectors retained), imaging factors (such as the number of cameras used, the image resolution, the number of frames per second for the video sequences, subject-to-camera distance, etc.), and variables such as illumination, orientation, scale, age, etc. As Table 8.1 clearly shows, the strategies mentioned in this paper widely differ in terms of these key factors, and, therefore,

Table 8.1. Summary of integrated face-gait systems

Authors	**Biometrics combined**	**Data**	**Recognition rates of classifiers**			
			Individual	**Fused**		
Zhou et al. [402] [405] [401] [34]	Face profile and gait silhouette [402]	14 people (2 outdoor video sequences each)	Face (64.3%), Gait (85.7%)	SUM rule (100%), PRODUCT rule (92.9%), Hierarchical method (92.9%)		
	Side face and gait silhouette [405] [401] [34]	45 people (2 to 3 video sequences each)	Face (Expt. 1 :91.1%, Expt. 2 :80%), Gait (Expt. 1 :93.3%, Expt. 2 : 82.2%)		**Expt. 1**	**Expt. 2**
				SUM rule [405]	95.6%	88.9%
				MAX rule [405]	97.8%	88.9%
				Feature fusion [401]	100%	88.9%
				Feature fusion [34]	100%	91.1%
Liu et al. [399]	Frontal face and gait silhouette	70 people (6 video sequences each, 6 static face images each)	Face (Expt. 1:40%, Expt. 2:40%), gait (Expt. 1:39%, Expt. 2:30%)		**Expt. 1**	**Expt. 2**
				SUM rule	71%	50%
				Bayesian rule	70%	50%
				Confidence weighted score sum	58%	55%
				Rank sum	68%	45%
Kale et al. [400]	Frontal face and gait silhouette	30 people (no. of video sequences not specified; static gallery images for the face)	Face (93%), Gait (61%)	SUM rule (100%), PRODUCT rule (100%), Hierarchical method (97%)		
Shakhnarovich et al. [33]	Frontal face and gait silhouette	12 people (2 to 6 VH sequences each)	Face (80%), Gait (87%)	Weighted sum (91%)		
Lee et al. [410]	Frontal face and trajectories of gait-defining points on body	12 people (2 video sequences each)	Not reported	Hierarchical method (Not reported)		



(a) (b)

Figure 8.1. Examples of face and gait representations used by: (a) Shakhnarovich et al. [33], (b) Zhou et al. [34]

do not lend themselves easily to efforts to analyze their performances relative to each other. However, a few general observations about these methods can be made based on the discussion above:

- All the techniques discussed above (excluding [410], for which the face classifier results are not specified) are reported to perform better than the individual face and gait classifiers, with the improvement in performance ranging from 5% to 35% for the face, and 2% to 39% for the gait. This indicates that combining the face and the gait cues for recognition is indeed beneficial.
- Most of the methods mentioned have examined the viability of combining these cues not only under optimal controlled conditions but also under difficult conditions such as changing viewpoint, aging or time delays, outdoors with uncontrolled lighting and at a distance. Their good performance even under these circumstances provides strong support for the practicability of deploying such systems in real-world scenarios and applications. In handling these variables, some techniques (such as those which utilize 3D information to cope with varying pose [33, 409]) may appear to suffer from the drawbacks of being computationally expensive and requiring large amount of memory as well as multiple cameras. However, with the continuing advances in processing power, computational capabilities and available data storage, it is reasonable to assume that these apparent shortcomings will become less of an issue in the future.
- All of the approaches examined have been tested on a small number of subjects so far. It remains to be seen if these systems will maintain their performance when tested on hundreds or thousands of individuals, as tends to be the case in several real-world application scenarios. Though large face image databases do exist for evaluating face

recognition systems [310], however, the largest gait database currently available, the NIST/USF Gait Challenge database [406], contains just 122 individuals. At its current state of development, computer-vision based gait recognition technology is generally not deemed mature enough to handle an inordinately large number of people but it does show promise as a screening device or filter for some other technology in such scenarios (similar to the way that it is being used in the hierarchical schemes mentioned above). If enough gait data is available, it would be interesting to see how techniques which directly combine face and gait at the feature or matching-score level fare in that situation.

- Though the initial forays into the domain of integrated face-gait recognition, exemplified by the strategies discussed in this paper, have yielded encouraging results, however, there are several issues which still need to be addressed: Additional work is required to cope with factors such as view invariance, aging, illumination variations, to name just a few, to enable such systems to handle the task of human recognition under unconstrained conditions with a reasonable degree of accuracy. Furthermore, only a few techniques, from the vast variety of strategies available for both face and gait recognition, have been incorporated into these systems. Alternative face and gait recognition methods need to be combined and empirically evaluated to determine which combination works best for different application scenarios.

CHAPTER 9

AN INTEGRATED FACE-GAIT SYSTEM FOR AUTOMATIC RECOGNITION OF HUMANS [6]

---

[6] Rabia Jafri, Hamid R. Arabnia and K. J. Simpson. Accepted and to appear in the Proceedings of the 2008 International Conference on Security and Management (SAM'08), July 14-17, Las Vegas, USA. Reprinted here with permission of publisher.

Abstract

A system that integrates the face, a physical biometric, with gait, a behavioral biometric, for automatically recognizing human beings is proposed. A decision-level fusion approach is adopted where the top matches of the face classifier are passed on to the gait classifier which then determines the identity of the unknown person. For gait recognition, a novel system is implemented, which utilizes various gait features identified as being the most pertinent for recognition based on data collected using an optoelectronic motion capture system.

Keywords: Face recognition, gait recognition, integrated face-gait recognition, person identification.

9.1. Introduction

Among the various biometric technologies explored in recent years, face recognition has emerged as one of the most promising biometric-based strategies to date with applications in domains such as surveillance, covert security and context-aware environments [37]. The face's appeal as a biometric stems from the several advantages that it offers in terms of being non-intrusive, non-invasive, cost-effective, easily accessible (i.e., face data can be conveniently acquired with a few inexpensive cameras) and relatively acceptable to the general public. Though current face recognition algorithms have reached a certain degree of maturity when operating under constrained conditions [37], however, there are some inherent weaknesses plaguing face recognition systems which diminish their effectiveness and essentially preclude their use in certain real-world situations (e.g., when the subject is too far away from the camera for his facial features to be clearly perceived, when the face is occluded or blurred or when the face is severely distorted due to extreme variations in expression, viewpoint, illumination, etc.) [413]. One obvious and feasible way to compensate for the degradation in the performance of the face recognition system in those scenarios is to use other biometric traits in conjunction with the face, which are not impeded by the same challenges as the face in those situations. Gait (a manner of walking or moving on foot) is a behavioral biometric characteristic which fits the bill.

The utility of gait for human recognition has only recently begun to be explored [39]. Similar to the face, gait, too, is a visual cue which can be extracted from video and thus, it appears to offer the same advantages (outlined above) that are presented by the face. However, gait offers some additional benefits in that, unlike the face, it can be acquired from a sequence of low resolution images of a person taken from a distance where the subject's body occupies too few pixels for other biometric traits to be discerned. Moreover, it is more robust to slight variations in viewpoint as compared to other biometrics and cannot be easily disguised without attracting attention. Nevertheless, gait as a biometric has its limitations since, being a behavioral trait, it can be affected and altered by factors such as clothing, footwear, environmental conditions, emotions, fatigue, drunkenness, pregnancy, injury, disease, aging and load.

In this paper, we propose combining the face, a physical biometric, with the gait, a behavioral biometric, with the aim of investigating if such a combination will improve upon the performance of the face classifier alone under circumstances where the face images available are of very low quality and resolution. A decision-level fusion approach is adopted where the top matches of the face classifier are passed on to the gait classifier which then determines the identity of the unknown person. For face recognition, the eigenfaces approach [6], as well as a Bayesian inference-based classifier from our previous work (described in [40]) is employed, while for gait recognition, a novel system is implemented, which utilizes the gait features identified as being the most pertinent for recognition based on data collected using an optoelectronic motion capture system.

The rest of the paper is organized as follows: Section 9.2 provides a brief overview of related work in the areas of face, gait and integrated face-gait recognition. Section 9.3 offers some additional insight into the motivation for combining the face and gait biometrics for human recognition. Section 9.4 describes the gait data collection process. Section 9.5 provides a description of the proposed system. Section 9.6 gives the details of the experiments conducted and the results obtained. Section 9.7 includes some concluding remarks and proposals for future work.

9.2. Related work

An overview of the methods currently being used for human recognition based on the face and gait, respectively, as well as by using a combination of these two biometric traits is given below.

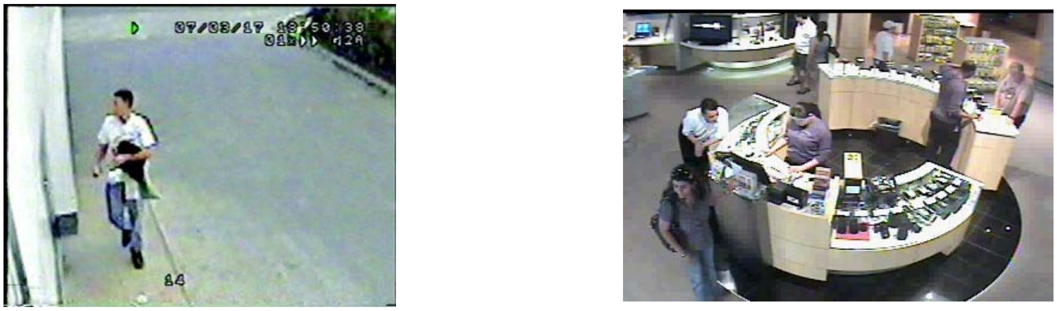
Face recognition has been the focus of active research for the past four decades (see [352] for a detailed survey). The approaches for this task can be broadly divided into two categories: 1) Feature-based methods [101, 102], which first process the input image to extract distinctive facial features, such as the eyes, mouth, nose, etc., as well as other fiducial marks and then compute the geometric relationships among those facial points, thus, reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed for matching faces using these measurements. 2) Appearance-based (or holistic) methods [6, 8, 414], which attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. Though face recognition methods traditionally operate on static intensity images, in recent years, much effort has also been directed towards identifying faces from video [254] as well as from other modalities such as 3-D [286] and infra-red [300].

Gait is a relatively new biometric characteristic. Methods to recognize gait from video sequences first emerged in the early 1990s mainly because it was around this time that computer memory and processing speed became sufficient to process large amounts of image data with reasonable efficiency [39]. Approaches that identify people from gait using video can broadly be partitioned into two groups [43]: 1) Model-based methods [42-44] use a model of either the person's shape or motion to derive static (i.e., absolute distances between certain body landmarks, such as the height, limb lengths, shoulder width, torso length, etc.) as well as dynamic gait features (e.g., swings of legs and arms, stride length, walking speed, and kinematics of joint angles) 2) Holistic techniques [389-391] represent the gait by the statistics of the spatiotemporal patterns generated by the walking person in an image sequence. It should be noted that some strategies have also been proposed to extract gait features from modalities other than video (e.g., Middleton et al. [392] utilize floor sensors to extract stride parameters as well as time on toe to time on heel ratios).

In recent years, some approaches have emerged which integrate the face and gait with the aim of investigating if such a combination will improve upon the performance of techniques which exclusively employ only one of these biometrics. A few of these methods are briefly described here: Zhou et al. [402] apply a curvature-based matching method [404] for classifying face profiles and direct matching of gait energy images derived from averaged gait silhouette sequences for classifying the gait. The face and gait classifiers are then combined using the SUM rule, the PRODUCT rule and a hierarchical scheme. Three variants of the above system are presented in later papers [34, 401, 405]. Liu et al. [399] integrate a frontal face classifier, based on the Elastic Bunch Graph Matching [102] algorithm, and a gait silhouette classifier, based on population Hidden Markov Models and eigenstance shape models, at both the decision level and the score level (using the SUM rule, Bayesian decision rule and Confidence Weighted Score Sum). Kale et al. [400] propose a view-invariant gait classifier and combine it with a still-to-video frontal face classifier [408] using a hierarchical approach as well as the SUM and PRODUCT rules. Shakhnarovich et al. [33] also perform view-invariant recognition by using 3-D visual hull to render frontal views for faces and side views for the gait. Faces are recognized using the eigenfaces approach [6], while gaits are classified based on the algorithm for matching spatiotemporal sequences proposed in [415]. Face and gait cues are then combined by taking a weighted sum of their confidence vectors. In a subsequent paper [409], this system was extended, with a more sophisticated probabilistic approach being used for combining the face and gait cues. A comprehensive review of integrated face-gait systems proposed so far can be found in [41].

9.3. Motivation for integrating face and gait

One major hurdle in employing face recognition in domains that can potentially benefit the most from this technology is the non-availability of high resolution and good quality face data in many application scenarios. For instance, Figures 9.1(a) and 9.1(b) show typical examples of faces seen in surveillance videos. Note that the faces occupy only a small portion of the whole image and are of much lower quality and resolution as compared to the face images usually utilized to test face recognition systems. Though previous studies [40, 48] have demonstrated that the performance of some well-known face recognition algorithms remains fairly stable with decreasing image resolution, with good recognition rates being achieved for resolutions as low as 64x64 pixels, however, these studies have also shown that when the image resolution is decreased even further, the performance of these algorithms degrades dramatically.



(a) (b)

Figure 9.1. Typical examples of images taken from surveillance videos [35]

Given the robustness of face recognition methods for resolutions as low as 64x64 pixels, recognition based on the face is a viable option even for applications using low-resolution face imagery. However, some mechanism is needed to compensate for the degradation in performance experienced by the face classifier when the image resolution falls below this threshold.

Since the performance of any classifier is more sensitive to some factors and relatively invariant to others, a recent trend has been to combine individual classifiers in order to integrate their complementary information and, therefore, create a system that is more robust than any individual classifier to variables that complicate the recognition task. Schiele et al. [397] have empirically demonstrated that as the number of classifiers combined increases, so does the recognition accuracy, while Hong et al. [396] have shown that integrating multiple biometrics does indeed result in consistent performance improvement. To this end, some combinations of the face with biometrics such as hand geometry [393], speech [315], fingerprint and speech [394], iris [395], and ear [74] have already been tested, with favorable results being reported. The positive outcomes reported in the literature for such systems [74, 315, 393-395] in conjunction with the conclusions reached by the studies mentioned above provide a strong basis for believing that a system constructed by combining various biometric characteristics is likely to yield better recognition rates than the individual classifiers for those traits. Moreover, exploiting more than one biometric for recognition will surely improve the reliability of the system and make it that much harder for someone to spoof it.

Therefore, one option for dealing with low resolution face data is to combine it with some other biometric characteristic. However, a major drawback of using any of the biometrics mentioned above is that acquiring data for them necessitates interaction with the subject, thus, compromising the advantages offered by the face in terms of being non-intrusive and non-invasive.

Gait, on the other hand, appears to be an ideal choice for a biometric to be used in conjunction with the face when the subject is so far away from the camera that only a low-quality and low-resolution image of his face can be acquired. As opposed to the face, a short-range biometric, gait, is a medium- to long-range biometric, that can be extracted reliably even from low-resolution imagery and is more invariant to slight changes in viewpoint. Also, both face and gait are visual cues, which can be extracted from the same modality, (i.e., image sequences of people) precluding the need for separate or specialized equipment. Furthermore, these biometrics make use of apparently independent personal characteristics: face recognition systems exploit the relatively detailed appearance of the facial surface, while gait recognition methods capture data from the coarse body shape as it changes over time. Consequently, some conditions that sharply degrade the performance of face recognition systems, such as large variations in illumination and facial expressions, affect gait to a much lesser extent or not at all. Similarly, some conditions that adversely affect the accuracy of gait recognition, such as clothing, footwear, and load, do not influence the performance of face recognition systems.

As mentioned in the previous section, the preliminary studies [41] exploring the combination of facial and gait cues for recognition have already revealed the potential of such an approach for discriminating among people. In light of the above discussion, it is reasonable to believe that utilizing gait information will enhance the performance of a face recognizer and will enable it to operate effectively in a wider range of conditions.

9.4. Gait data collection

For this study, the spatial locations of body points were collected using a seven-camera visible-red light optoelectronic motion capture system (Vicon MX-40®, Vicon Motion Systems, Oxford, UK: 40 megapixel cameras, sample rate = 120 frames per second).

24 healthy participants (12 male, 12 female, height: 170.9±8.17 cm, mass: 67.742±14.21 kg, age: 23.0±3.9 yr) with no known illnesses or injuries that could affect gait performance were recruited. Each potential participant signed an informed consent form as approved by our institutional review board, then completed a health status and demographics (e.g., age, limb dominance) questionnaire.

Prior to data collection, anthropometric measures (height, mass, leg lengths and knee and ankle joint widths) were obtained. Thirty-nine 14 mm reflective markers were then attached to the skin or clothing in accordance with the Vicon Plug-in-Gait® model [416] that is based on the modified Helen-Hayes model [417] (Table 9.1 and Figure 9.2). To later establish the spatial relationships among the markers, the segmental coordinate systems, and room coordinate system, a static trial of the person standing in a natural position was captured. Next, the participant practiced the walking task three to five times.

For the testing, the participant performed 15 acceptable trials. For one trial, the participant walked approximately 10 m, while stepping onto an AMTI® force platform (1200 Hz) mounted flush to the floor at approximately the midpoint of the walkway. A trial was acceptable if the participant did not self-report making any gait adjustments to strike the platform, the

Table 9.1. Reflective marker placement. The abbreviations below are used in the accompanying diagram (Figure 9.2)

Marker name	Abbreviation	Location
Left and right front head	LFHD, RFHD	Over each temple
Left and right back head	LBHD, RBHD	Back of the head, roughly in a horizontal plane of front head markers
7th cervical vertebra	C7	Spinous process of the 7th cervical vertebra
10th thoracic vertebra	T10	Spinous process of the 10th thoracic vertebra
Clavicle	CLAV	Jugular notch where the clavicles meet the sternum
Sternum	STRN	Xiphoid process of the sternum
Right back	RBAK	Middle of the right scapula
Left and right shoulders	LSHO, RSHO	Each acromio-clavicular joint
Left and right upper arms	LUPA, RUPA	Each upper arm between elbow and shoulder markers
Left and right elbows	LELB, RELB	Each lateral epicondyle approximating elbow joint axis
Left and right forearms	LFRA, RFRA	Each lower arm between the wrist and elbow markers
Left and right wrist A	LWRA, RWRA	Each wrist distal radius
Left and right wrist B	LWRB, RWRB	Each wrist distal ulna
Left and right fingers	LFIN, RFIN	Dorsum of each hand just below the head of the 2nd metacarpal
Left and right ASIS	LASIS, RASIS	Each anterior superior iliac spine
Left and right PSIS	LPSI, RPSI	Each posterior superior iliac spine
Left and right knees	LKNE, RKNE	Lateral epicondyle of each knee
Left and right thighs	LTHI, RTHI	Lower lateral 1/3 surface of each thigh
Left and right ankles	LANK, RANK	Center of each lateral malleolus
Left and right tibia	LTIB, RTIB	Lower 1/3 of each shank
Left and right toes	LTOE, RTOE	2nd metatarsal head
Left and right heels	LHEE, RHEE	Each calcaneous

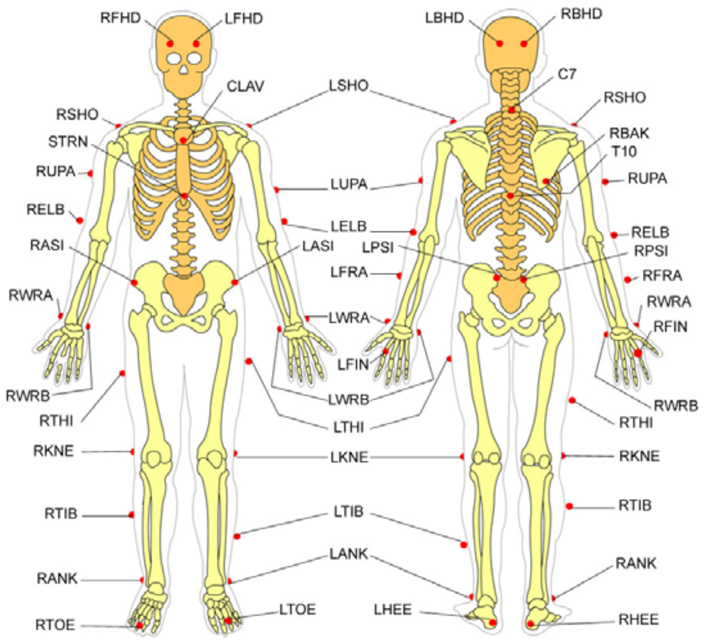


Figure 9.2. Marker placement for the Plug-in-Gait® model [36]

researchers did not visually observe such adjustments and observed a typical vertical ground reaction force pattern, and the subject hit the force platform with the foot entirely on the platform without making adjustments to the gait and maintained a constant walking speed.

9.4.1. Processing the gait data

The 3D spatial locations of the markers were reconstructed (Plug-in-Gait® Vicon Workstation® software), then smoothed using a generalized, cross-validatory spline [418]. The data from the static trial were used to construct the transformation matrices needed to generate the segmental coordinate systems of the trunk and lower extremity segments relative to the room coordinate system for the gait trials. The stride analyzed started at heel strike onto the force platform and ended with the next heel strike of the same foot. Overall gait kinematics (e.g., step length), joint kinematics (e.g., joint angle trajectories) and kinetics (e.g., forces, moments) were generated (Plug-in-Gait® module). The data were scaled temporally to the total stride time (ST) in increments of 1% ST using a 4-point polynomial interpolative method [419].

9.5. Integrated face-gait recognition system

A description of the classifiers employed for recognizing the face and gait, along with an explanation of the approach used to combine these classifiers to form an integrated face-gait recognition system, is provided in this section.

9.5.1. Face classifier

Two different appearance-based approaches are employed for classifying faces: the PCA-based eigenfaces method [6] and the Bayesian inference-based technique presented in our previous work [40] which combines a maximum a posteriori (MAP) classifier with a maximum likelihood (ML) classifier. The first is a well-established method in face recognition literature, frequently used as a baseline for evaluating face recognition algorithms [84], that offers the advantage of speed and simplicity while yielding reasonably accurate results under controlled conditions. The second strategy was selected since our previous experiments indicated that this technique produces the best recognition rates for low-resolution face images. This method, though significantly more accurate than the eigenfaces technique, suffers from the drawback of being computationally intensive.

The code for the eigenfaces, MAP and ML classifiers was acquired from the CSU face identification evaluation system [312]. The implementation details may be found in the user's guide for this system [313]. Details on combining the MAP and ML classifiers may be found in [40].

9.5.2. Gait classifier

9.5.2.1. Motivation for using a model-based approach. We decided to apply a model-based approach, as opposed to a holistic one, based on the following considerations: Holistic methods offer the advantages of speed and simplicity, but have the major disadvantage that spatio-temporal changes in the whole silhouette are only indirectly linked to gait dynamics. Hence, it is hard to infer the importance of various gait components from the moving silhouette and it is not clear how to normalize the silhouette data to reduce the effects of noise, clothing and other dependent factors [420]. Model-based approaches, on the other hand, though more computationally complex, take knowledge of the shape and dynamics of human gait into account while extracting features, and gait dynamics are calculated based on anatomical landmarks rather than changes in the overall silhouette. Thus, these techniques are more immune to the effects of clothing as well as slight changes in viewpoint [421]. Moreover, they allow us to determine the relative importance of various gait components for recognition.

9.5.2.2. Motivation for acquiring gait information from motion capture data. We decided to utilize gait features extracted from motion capture data, rather than video, in our approach. The decision was motivated by the need to avoid inaccuracies in gait data stemming from the usual problems of extracting visual cues from video such as imperfect foreground segmentation of the walking subject from the background scene and poor imaging conditions. Obviously, in a real-world scenario, it would be much more practical to acquire gait data from a video sequence. However, our goal at this stage is to come up with methods to extract identity information present in gait features and for this purpose, we want to consider the ideal case where the gait data available is as accurate and complete as possible. Moreover, this might lead to the discovery of some new features, useful for discrimination, which may not be easily recovered from video alone, but which could be obtained using other means. Current gait recognition methods (all of which, excepting a few (e.g., [422]), are video-based) have already demonstrated that certain gait information can be extracted from video with some degree of reliability and as vision systems become more robust at recovering gait data from video, it is reasonable to expect that the results obtained from video data would reflect the ones presented here based on motion capture data.

9.5.2.3. Forming the gait signature. Gait features currently utilized by model-based methods can be broadly divided into two categories [422]: 1) Static features depend upon geometry based measurements, such as distances between anatomical landmarks (e.g., height [43], torso length, limb lengths [44]). 2) Dynamic features are dependent on the motion of the body over time. These include general motion parameters such as walking speed [423], stride time[423], cadence (number of steps per minute) [43] as well as kinematic variables (i.e., variables involved in the geometric description of movement, independent of forces that cause that movement [424]) such as linear and angular displacements, velocities, accelerations [423] and joint angle trajectories [422].

Since static parameters are highly dependent on the subject's clothing, the practical utility of these variables for recognition is debatable. Therefore, in our preliminary investigation, we have incorporated only dynamic parameters in our characterization of gait. The following spatio-temporal parameters were selected: 1) stride length (distance between 2 successive heel strikes of the same foot), 2) stride time (time for one stride), 3) step length (distance between the heel strike of one foot to the heel strike of the opposite foot), 4) step time (time for one step), 5) step width (lateral distance between both heel centers of 2 consecutive foot contacts), 6) cadence (number of steps per minute), 7) walking speed (distance per unit of time), 8) stance phase time (the duration when the foot is in contact with the ground, i.e., heel strike to toe off), 9) swing phase time (the duration when the foot is in the air, i.e., toe off to heel strike).

Also, the trajectories of the following angles, about all three axes (Plug-in-Gait®) were chosen [36]: 1) trunk: thorax, spine, neck and pelvis angles; and for both sides of the body, 2) shoulder and elbow joint angles 3) ankle, knee and hip joint angles and 4) foot progression angle (angle of foot relative to the direction of walking progression).

The gait signature is created simply by concatenating these parameters to form a gait feature vector. Section 9.6.3.1 describes in detail the results obtained by using different combinations of these parameters.

9.5.2.4. Recognizing gait. In the following discussion, using terminology borrowed from face recognition [84], a gallery refers to a set of known individuals while a probe refers to an unknown person.

Let $x_0, \ldots, x_m$ be spatiotemporal variables. Let $t_0, \ldots, t_n$ be angle trajectories, where $t_i=[t_{i0}, t_{i1}, \ldots, t_{i100}]$, $i=0, \ldots, n$. Let SED be the squared Euclidean distance between two vectors of equal length. Then, given two vectors $v_1=[v_{10}, \ldots\ v_{1k}]$, $v_2=[v_{20}, \ldots, v_{2k}]$, SED($v_1$, v_2) = $\sum_{i=0,\ldots,k}(v_{1i}-v_{2i})^2$.

Let $gait_1, \ldots, gait_g$ be the gallery gait feature vectors, where $gait_i=[x_{i0}, x_{i1}, \ldots, x_{im}, t_{i0}, t_{i1}, \ldots, t_{in}]$, $i=1, .., g$. Given a probe gait feature vector $gait_p=[x_{p0}, x_{p1}, \ldots, x_{pm}, t_{p0}, t_{p1}, \ldots, t_{pn}]$, the similarity scores between the probe and the gallery vectors are calculated as follows: For each gallery vector $gait_i$: 1) The squared distances between each corresponding pair of spatiotemporal variables, $(x_{ij}-x_{pj})^2$, $j=0, \ldots, m$, are computed. 2) The squared Euclidean distances between each corresponding pair of angle trajectories, SED(t_{ij}, t_{pj}), $j=0, \ldots, n$, are computed.

To ensure that each variable is given equal weight in the final similarity score, the distances computed in 1 and 2 are normalized between 0 and 1 as follows: For a particular variable, let $d_{1p}, \ldots, d_{gp}$ be the distances between the g gallery gaits and the probe p. Then, for gallery i, the normalized distance for this variable is computed as follows:

$$norm(d_{ip}) = (d_{ip} - min(d_{1p}, \ldots, d_{gp}))/(max(d_{1p}, \ldots, d_{gp}) - min(d_{1p}, \ldots, d_{gp})) \quad \textbf{(9.1)}$$

The final similarity score between gallery gait i and probe gait p is calculated as follows:

$$D(gait_i, gait_p) = \sqrt{\sum_{j=0,\ldots,m} norm((x_{ij} - x_{pj})^2) + \sum_{j=0,..,n} norm(SED(t_{ij}, t_{pj}))} \quad \textbf{(9.2)}$$

The gallery gaits are sorted in decreasing order of their similarity scores. The probe is considered to be correctly recognized if it matches the first gallery gait on the sorted list.

9.5.3. Integrated face-gait recognition system

We first discuss various general schemes used to fuse biometrics before describing the approach selected by us.

Fusion of biometrics can be done at three levels [393]: 1) The feature level, where feature vectors of the individual biometrics are concatenated to form a new feature vector, which represents the subject's identity in a higher-dimensional and, presumably, more discriminating hyperspace. 2) The matching score level, where the similarity scores of the individual biometrics are first transformed (using transformations such as linear, exponential, logarithmic, etc.) to ensure that they map to a common range and then combined using some combination rule (e.g., SUM, PRODUCT, MAX, etc.) to obtain a new similarity score. 3) The decision level, where each individual classifier makes its own decision regarding the identity of the unknown person and some rule (e.g., majority voting, rank rule, etc.) is employed to make the final decision.

Though most of the integrated face-gait systems proposed so far make use of decision-level and matching-level fusion [399] [400], nevertheless, a couple of methods also utilize feature-level fusion [401] [34].

We decided to employ a decision-level fusion approach to integrate our face and gait classifiers. When presented with an unknown person (probe), the system first runs the face classifier to classify the person's face. A certain percentage of the closest matches of the face classifier are then passed on to the gait classifier which then chooses the best match for the probe from this smaller set based on gait.

9.6. Experiments

9.6.1. Face and gait data acquisition

Face data for these experiments was taken from the FERET database [310]. The frontal images in this database were extracted and from these, the images of subjects wearing glasses were manually selected and discarded (since it has been shown that glasses make subjects more recognizable [311]). The resulting collection of images contained 3203 images of 1127 subjects. A training set was then constructed from the resulting data set by selecting subjects with exactly three images. 203 such subjects (and hence, 609 images) were found. The remaining 2594 images of 924 subjects were declared as potential candidates for inclusion in probe and gallery sets.

The face gallery and probe sets were then constructed as follows: 12 male and 12 female subjects were randomly selected from the 924 available subjects. Two different face images were selected for each of these subjects: one image was placed in the gallery set while the other was added to the probe set.

The gait data, described in section 9.4, were used. A main gait gallery set with 24 gaits was constructed by randomly selecting one trial for each subject. A corresponding main probe set with 24 gaits was then built by randomly selecting one trial for each subject such that that trial was different from the one included in the gallery set. Smaller gallery sets of sizes 10, 15 and 20 were similarly constructed by randomly selecting trials from the main gallery set such that no two trials in each set belonged to the same subject. The corresponding probe sets of the same sizes were constructed by selecting one trial from the main probe set of each subject in the gallery set such that that trial was different from the one included in the gallery set.

Each male face subject was associated with a unique male gait subject. Similarly, each female face subject was associated with a unique female gait subject.

9.6.2. Preprocessing of data and similarity score calculation

The resolution of the face images was reduced to 16x24 pixels. These low-resolution face images were then normalized using the code provided in the CSU face identification evaluation system [312]. Training was conducted for the eigenfaces, Bayesian-MAP and Bayesian-ML classifiers as described in [40]. The energy criteria proposed by Kirby [314] was selected for deciding the number of subspace basis vectors to retain and a threshold of 0.9 was chosen. The gallery and probe images were then projected onto their respective subspaces and the distances/similarity scores among them were calculated as explained in [40].

The gallery and probe gaits were processed as described in section 9.4. In the sub-sections following this one, the gait feature vectors for the gallery and probe gaits are formed and similarity scores for them are calculated as described in sections 9.5.2.3 and 9.5.2.4, respectively.

9.6.3. Tests

9.6.3.1. Gait classifier tests.

9.6.3.1.1. Tests with spatio-temporal parameters. The recognition potential of the spatio-temporal parameters was studied by running the gait classifier using these parameters individually, as well as the combination of them all. The results are depicted in Table 9.2. Various combinations of the parameters with the highest recognition rates were then taken. The results are shown in Table 9.3.

Table 9.2. Gait classifier recognition rates using spatio-temporal parameters

Gallery Size	Stride length	Stride time	Step length	Step time	Step width	Cadence	Walking speed	Stance phase time	Swing phase time	All spatio-temporal parameters
10	40.00%	50.00%	30.00%	90.00%	0.00%	90.00%	40.00%	80.00%	30.00%	90.00%
15	40.00%	20.00%	26.67%	73.33%	13.33%	73.33%	20.00%	60.00%	26.67%	100.00%
20	20.00%	30.00%	25.00%	60.00%	15.00%	55.00%	25.00%	40.00%	25.00%	90.00%
24	25.00%	25.00%	29.17%	62.50%	12.50%	58.33%	20.83%	29.17%	20.83%	79.17%

Table 9.3. Gait classifier recognition rates using various combinations of spatio-temporal parameters

Gallery Size	Step time + cadence	Step time + cadence + stance phase time
10	90.00%	100.00%
15	73.33%	86.67%
20	55.00%	75.00%
24	58.33%	66.67%

The following observations can be made based on these results:

- Recognition rates decrease in general as the gallery size increases. The recognition rates based on individual parameters are lower than those based on the combination of all the parameters (except for the smallest gallery size).
- When the two best-forming parameters (i.e., step time and cadence) are combined, the recognition rates are equal to or worse than those based on individual parameters. They are also worse than those based on the combination of all the parameters.
- When the three best-performing parameters (i.e., step time, cadence, and stance phase time) are combined, the recognition rates are better than for the above combination but worse than those based on the combination of all the parameters (except for the smallest gallery size).

The above observations indicate that the low-performing parameters are contributing discriminatory information which complements that contributed by the higher-performing variables.

9.6.3.1.2. Tests with angle trajectories. The recognition potential of the angle trajectory parameters was studied by running the gait classifier using these parameters individually. The results for some of the better-performing variables are depicted in Table 9.4. The recognition rates obtained by taking various combinations of these parameters are shown in Table 9.5. Note that to ensure that the results presented here reflect real-world conditions more closely (where the gait data will be extracted from video and thus, will be much less accurate than that being utilized in

Table 9.4. Gait classifier recognition rates using angle trajectories

Gallery Size	Left hip angles			Right hip angles			Left knee angles			Left ankle angles		
	1	2	3	1	2	3	1	2	3	1	2	3
10	100%	100%	100%	90.00%	100%	100%	100%	100%	100%	100%	100%	100%
15	80.00%	93.33%	93.33%	80.00%	100%	100%	80.00%	93.33%	100%	86.67%	93.33%	86.67%
20	90.00%	90.00%	95.00%	85.00%	95.00%	100%	85.00%	95.00%	95.00%	95.00%	95.00%	85.00%
24	79.17%	91.67%	95.83%	75.00%	95.83%	100%	87.50%	95.83%	95.83%	91.67%	95.83%	87.50%

Table 9.5. Gait classifier recognition rates using various combinations of angle trajectories (rr = individual recognition rates of angle trajectories being combined for largest gallery size)

Gallery Size	All	Lower body	Upper body	rr>=87.5%	rr<=75%	rr<=66.67%	rr<=62.5%	rr<=54.16%	rr<=45.83%
10	90.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	90.00%	80.00%
15	100.00%	100.00%	100.00%	93.33%	100.00%	100.00%	100.00%	93.33%	93.33%
20	90.00%	95.00%	90.00%	95.00%	90.00%	90.00%	90.00%	85.00%	70.00%
24	79.17%	95.83%	91.67%	95.83%	91.67%	91.67%	91.67%	79.17%	70.83%

this study), twice the weight has been given to some low-performing variables, i.e., the shoulder, neck, spine and pelvis angles, when they are present in a combination.

The following observations can be made based on these results:

- The recognition rates based on the combination of all the parameters are less than those based on some of the individual parameters. This is not surprising since a combination of a large number of parameters is being taken, many of which are not yielding good results individually, and the discriminatory potential of the better-performing variables is nullified by the several low-performing ones.
- Component 3 of the right hip angle trajectory is the best performing variable overall, yielding 100% recognition rates for all gallery sizes. This might indicate that this angle alone is sufficient for classifying people based on their gait. However, it should be noted that these rates are being obtained using highly accurate data for this angle and for a

small number of individuals. It would be unrealistic to expect such ideal rates in a real-world scenario with more noisy data and a larger number of subjects. Moreover, in practice, recognition would have to be performed based on the subset of angles which are available for a certain individual and that subset might not include this particular angle. So despite the perfect recognition rates, it would not be prudent to rely exclusively on this angle to characterize the gait.

- When only the lower body angles are combined, the recognition rates are better than those obtained for the combination of all parameters (except for gallery size 15, where the rate is the same).
- When only the upper body angles are combined, the recognition rates are equal to or better than those for the combination of all parameters. They are also equal to those for the lower body angles for the first two gallery sizes but worse than those for the lower body angles for the last two gallery sizes. This indicates that the lower body angles are more useful for recognition than the upper body angles for larger gallery sizes.
- When the angles yielding recognition rates >= 87.5 % for the largest gallery size are combined, the resulting rates are the same as those for the combination of all lower body angles (except for gallery size 15 where the rate is slightly less). This implies that it might suffice to use this set of angles to attain the same recognition performance as that obtained by using all the lower body angles.
- When the angles yielding recognition rates <=75% for the largest gallery size are combined, the rates are equal to the combinations of all lower body angles for the two smallest gallery sizes and only slightly less than the combination of all lower body angles for the two larger gallery sizes. As combinations of angles with progressively less

recognition rates (i.e., 66.67%, 62.5%) are considered, the resulting recognition rates remain constant (i.e., the same as those for the combination of angles with rates <= 75%). This indicates that even variables that yield low recognition rates individually have enough discriminatory power so that when combined, they are capable of yielding much higher recognition rates. At 54%, the recognition rates for all gallery sizes fall slightly. Finally, at 45%, there is a decrease of 20% in the recognition rates for all gallery sizes (except gallery size=15, for which the decrease is 7%).

9.6.3.1.3. Tests with combinations of spatio-temporal parameters with angle trajectories.

The results in sections 9.6.3.1.1 and 9.6.3.1.2 show that the best recognition rates based on spatio-temporal parameters are obtained by using the combination of all these parameters and the best recognition rates based on combinations of angle trajectories are obtained by using a combination of all the lower body angles.

A combination of all spatio-temporal parameters and all lower body angle trajectories was, therefore, taken to explore if such a combination would yield better recognition rates. However, as shown in Table 9.6, this yields recognition rates equal to those obtained by the combination of all lower body angle trajectories alone. However, combining all spatio-temporal

Table 9.6. Gait classifier recognition rates using various combinations of spatio-temporal parameters and angle trajectories

Gallery Size	All spatio-temporal parameters + all lower body angles	All spatio-temporal parameters + all angles
10	100.00%	100.00%
15	100.00%	100.00%
20	95.00%	95.00%
24	95.83%	95.83%

parameters with all angle trajectories produces higher recognition rates (shown in Table 9.6) than using a combination of either of these classes of variables alone.

9.6.3.2. Integrated face-gait classifier tests.

Since it is reasonable to assume that data for certain gait parameters would be available in real-world application scenarios and not for others, therefore, three different gait classifiers, utilizing the following three sets of variables, respectively, were constructed to evaluate the proposed system: 1) all spatio-temporal variables, 2) all angle trajectories with recognition rates >= 87.5% for gallery size = 24, 3) all angle trajectories with recognition rates <= 54.16% for gallery size = 24. For each gait classifier, two integrated face-gait classifiers were constructed by combining the eigenfaces-based face classifier and the Bayesian inference-based face classifier (described in section 9.5.1), respectively, with the gait classifier. These integrated classifiers were tested using the gallery and probe sets containing all 24 subjects. For both of these integrated systems, the recognition rates using the face alone, the gait alone and the face and gait together are shown in Table 9.7. It should be noted that since both integrated systems use the same gait classifier,

Table 9.7. Recognition rates using face, gait and face+gait when gait classifiers 1, 2 and 3, respectively are utilized. "Percentage passed" refers to the percentage of closest matches of the face classifier that are passed to the gait classifier

Gait classifier utilized		Face	Gait	Face+gait		
	Percentage passed			30%	20%	15%
1	System 1	79.17%	79.17%	95.83%	100%	95.83%
	System 2	87.50%	79.17%	95.83%	100%	100%
2	System 1	79.17%	95.83%	100%	100%	95.83%
	System 2	87.50%	95.83%	100%	100%	100%
3	System 1	79.17%	79.17%	83.33%	95.83%	95.83%
	System 2	87.50%	79.17%	83.33%	91.70%	100%

therefore, any difference between their performances directly follows from the difference in the performances of their face classifiers.

The following observations about the systems utilizing gait classifier 1 can be made based on these results:

1) Both integrated face-gait classifiers yield better recognition rates than the individual face and gait classifiers.
2) The improvement in performance relative to the face is ~16 to 20% for the first integrated classifier. This shows that adding gait information to the eigenfaces classifier does significantly enhance its performance. In particular, note that the eigenfaces classifier alone was producing much lower classification rates (~8 % less) than the more powerful Bayesian face classifier. However, integration with the gait classifier enables it not merely to equal but to exceed the performance of the Bayesian classifier.
3) The improvement in performance relative to the face is ~8 to 12% for the second integrated classifier. This shows that the gait information does aid the face classifier in this case, too, albeit to a lesser extent as compared to the eigenfaces classifier.
4) The improvement in performance relative to the gait classifier is the same for both systems (~16 to 20%).
5) The highest recognition rates attained by both integrated classifiers are the same.
6) From observations 2, 3 and 5, it follows that a simple and fast face classifier combined with gait can achieve the same performance as a complex and slow face classifier combined with gait. This may have practical implications for applications seeking recognition at a faster speed without compromising accuracy.

The above observations also hold for the systems utilizing gait classifier 2, with the following exceptions:

1) The first integrated system yields recognition rates equal to (rather than better than) those obtained by the gait classifier alone when 15% of the closest matches of the face classifier are passed to the gait classifier.
2) The improvement in performance relative to the gait classifier is 0-4% for both systems.

The first three observations about the systems utilizing gait classifier 1 also hold for the systems utilizing gait classifier 3 with the following exceptions:

1) The second integrated system yields recognition rates worse than those obtained by the face classifier alone when 30% of the closest matches of the face classifier are passed to the gait classifier.
2) The improvement in performance relative to the face is ~4 to 16% for the first integrated classifier and ~4 to 12% for the second one.
3) The improvement in performance relative to the gait classifier is ~4 to 16% for the first integrated classifier and ~4 to 20% for the second one.

It follows from the above that the integrated face-gait systems yield better recognition rates than those obtained by the individual face and gait classifiers not only when the best-performing gait parameters are selected but also when low-performing gait variables are combined. This demonstrates the real-world application potential of the proposed system and the practicability of employing such an approach even in scenarios where a limited number of gait parameters are available (and even if they are ones that do not perform well individually). That being said, it should be noted that even combinations of low-performing variables are yielding high recognition rates for small gallery sizes. Since the present data set contains only 24 subjects, hence, the small percentage of that set passed to the gait classifier consists of only a few individuals. Thus, as long as the face classifier places the correct match in that set, even a weak gait classifier would be likely to recognize it. A clearer picture of the performance of this system will emerge when it is evaluated using a larger number of subjects.

9.7. Conclusions and future work

A system that integrates the face with the gait for automatically recognizing human beings was presented in this paper. Two different face classifiers operating on low resolution face images were separately combined with various gait classifiers which utilize dynamic gait features extracted from motion capture data. In all cases, the integrated system outperformed the individual face and gait classifiers that it was composed of, thus, demonstrating the potential of using the gait to supplement the face in scenarios where the face classifier alone does not perform well due to the non-availability of high resolution face data.

We plan to continue developing this system to enhance its performance and intend to more thoroughly investigate its potential for use in real-world application scenarios. Some directions being considered for future work are outlined below:

- Due to the time-consuming nature of the motion data collection process, gait data for only a small number of subjects had been collected at the time this study was conducted. The system has, thus, currently been tested with 24 subjects. The gait data collection is still underway, and as gait data for more subjects becomes available, the system's performance for larger data sets will be evaluated. Some other options for dealing with the scarcity of gait data at this point include generating synthetic gaits from the available gaits by combining them and applying the confusion metric introduced in [425] to predict the system's performance on larger populations.

- More sophisticated techniques, such as those presented in [426], may be utilized for selecting the best combination of gait features for recognition. Another option is to determine the potency of each feature for discrimination and to assign weights to the features accordingly (e.g., the F-statistic is used to rank and assign weights to various features in [420]).
- Additional gait features, such as linear and angular velocities and accelerations and correlations between angles, may be used for recognition.
- A simple nearest-neighbor classifier based on Euclidean distances is presently being used for gait classification. Other classification methods and distance measures [309] also need to be explored for recognizing the gait.
- Other strategies for fusing the face and gait, such as feature-level combination of these traits, may be investigated.
- As mentioned before, even though gait information for this study was obtained from motion capture data, our goal, ultimately, is to extract this information from video sequences. One important step in this direction is to examine how sensitive the data is to slight changes in the marker positions. For this study, the marker positions for a particular subject did not change from trial to trial. However, it is reasonable to believe that when these anatomical landmarks are located in video, the positions determined for them will vary slightly from one trial to another. Therefore, it is important to examine how gait data is affected by minor variations in the positions of these points. To this end, motion capture data for some of the subjects will be collected again. Since the marker positions

will be slightly different from the ones in the previous session, they would simulate the positions extracted from video. The similarity of the gait data to that collected in the previous session would be evaluated. Such a study would aid in assessing the applicability of the results shown here to video data.

- Since information about variables such as age, gender, limb dominance and emotional state, etc. is being collected for all gait participants, the effects of these variables on gait recognition can also be eventually studied. It should be noted that some studies have already been conducted to examine how these factors relate to gait (e.g., [427, 428] found that an individual's gender can be identified based on gait).

Acknowledgements - We are grateful to Sabrina Fordham and Yang-Chieh Fu from the Department of Kinesiology and Syed Abid Ali from the Department of Biochemistry at the University of Georgia for assisting in the gait data collection and the motion capture system set up. We would also like to thank the students from the Department of Computer Science and the Department of Kinesiology at the University of Georgia who volunteered to participate in this study.

CHAPTER 10

EVALUATING THE INTEGRATED FACE-GAIT RECOGNITION SYSTEM WITH A LARGER DATA SET

10.1. Introduction

As mentioned in the previous chapter, due to the time-consuming nature of the gait data collection process using an optoelectronic motion capture system, gait data for only a small number of subjects has been collected to date. The integrated face-gait recognition system presented in the previous chapter has, therefore, been tested with only 24 subjects.

However, the real utility of the proposed system lies in scenarios where hundreds of individuals need to be distinguished from each other. Therefore, in order to realistically evaluate the application potential of the integrated face-gait system and to examine the scalability of this approach to larger data sets, the system needs to be tested with a much larger number of individuals.

One option for dealing with the scarcity of gait data at this stage is to generate synthetic gait data by combining the gait data collected so far. In this chapter, therefore, we describe how such synthetic gait data can be computed and then associated with face images to generate synthetic individuals. The results obtained by testing the gait recognition and integrated face-gait recognition systems with this larger data set are reported and an analysis of how the performance of these systems with this larger set compares to that using a set of only 24 individuals is provided.

The rest of this chapter is organized as follows: Section 10.2 describes how synthetic gait vectors are computed from real gait vectors and then associated with face images to generate synthetic individuals. Section 10.3 reports the results obtained by empirically testing the gait recognition and integrated face-gait systems with the large data set consisting of real and synthetic individuals and analyzes how the recognition accuracy of these systems with this set compares to that using the smaller data set of 24 real individuals. Section 10.4 concludes the chapter.

10.2. Generating synthetic individuals

Suppose that gait data has been collected for n individuals using an optoelectronic motion capture system, with k trials being captured for each subject as described in section 9.4 in the previous chapter. Let $g_1, \dots, g_k$ be the gait vectors constructed for each individual (as described in section 9.5.2.3 in the previous chapter).

The gait data for a synthetic individual can then be constructed from the gait data for individuals i and j simply by taking the averages of their gait vectors. More specifically, if $g_{i1}, \dots, g_{ik}$ and $g_{j1}, \dots, g_{jk}$ are the gait vectors for individuals i and j, respectively, then the gait vectors for the synthetic individual generated from individuals i and j would be $(g_{i1}+g_{j1})/2, \dots, (g_{ik}+g_{jk})/2$.

Since previous studies have shown that a person’s gait is dependent on his/her gender [427, 428], hence, based on this consideration, our original data set contains an equal number of male and female participants. We continue to take gender into account when generating synthetic individuals by combining female gaits only to generate gaits for synthetic female individuals and combining male gaits only to generate gaits for synthetic male individuals.

The total number of synthetic individuals that can be generated from a set of n individuals using the above approach can then be computed as follows: Let n_f and n_m be the number of female and male subjects respectively, such that $n = n_f+n_m$. Since the synthetic individuals are being generated by combining two individuals, hence, the number of synthetic female individuals n_{sf} that can be generated from the n_f female subjects is equal to the number of unique combinations of n_f subjects taken two at a time:

$$n_{sf} = \frac{n_f!}{(n_f - 2)!2!} = \frac{n_f(n_f - 1)}{2} \qquad \textbf{(10.1)}$$

Similarly, the number of synthetic male individuals that can be generated from the n_m male subjects is equal to the number of unique combinations of n_m subjects taken two at a time:

$$n_{sm} = \frac{n_m(n_m - 1)}{2} \qquad \textbf{(10.2)}$$

Hence, the total number of synthetic individuals that can be generated from n subjects is

$$n_s = n_{sf}+n_{sm} \qquad \textbf{(10.3)}$$

Each synthetic female gait can then be associated with a unique female subject and each synthetic male gait can be associated with a unique male subject from the face database.

10.3. Experiments

10.3.1. Face and gait data acquisition, preprocessing and similarity score calculation

Gait data for 24 subjects (12 female, 12 male) was collected and processed as described in section 9.4 in the previous chapter. The gait feature vectors for these subjects were formed as described in section 9.5.2.3 in the previous chapter.

Since the number of female subjects, $n_f = 12$ and the number of male subjects, $n_m = 12$, substituting these values in equations (10.1) and (10.2), the following values can be calculated:

The number of synthetic females, $n_{sf} = 12(12-1)/2 = 66$

The number of synthetic males, $n_{sm} = 12(12-1)/2 = 66$

Substituting the values of n_{sf} and n_{sm} in equation (10.3):

The total number of synthetic individuals, $n_s = n_{sf} + n_{sm} = 132$

Hence, the total number of individuals (both real and synthetic) in the new data set = $n + n_s = 24 + 132 = 156$.

A gait gallery set with 24 gaits was constructed by randomly selecting one trial for each real subject. A corresponding gait probe set with 24 gaits was then built by randomly selecting one trial for each real subject such that that trial was different from the one included in the gallery set. The gallery gait for each synthetic individual was formed by averaging the gallery gait vectors of its constituent real individuals and was then added to the gallery set. Similarly, the probe gait for each synthetic individual was formed by averaging the probe gait vectors of its constituent real individuals and was added to the probe set. The resulting gallery and probe sets contained 156 gaits each, one for each real and synthetic individual.

The face data was acquired, preprocessed, divided into training, gallery and probe sets and the similarity scores among the gallery and probe images were calculated as described in sections 9.6.1 and 9.6.2 in the previous chapter. However, note that 78 (instead of 12) unique female subjects and 78 (instead of 12) male subjects were now selected from the 924 available subjects.

Each female gait subject was associated with a unique male face subject. Similarly, each male gait subject was associated with a unique male face subject.

10.3.2. Tests to evaluate gait recognition and integrated face-gait recognition systems

In section 9.6 of the previous chapter, various tests to evaluate the gait classifier and the integrated face-gait system were described and the results for those tests using gallery and probe sets of 24 or fewer individuals were reported. These tests were conducted again, this time using gallery and probe sets of 156 individuals (constructed as described in the previous section) and the results generated are reported in the following subsections. At the end of each subsection, a comparison of the results using this larger set with the results obtained previously using 24 subjects is provided. (Note that for each test, the corresponding results previously obtained using 24 individuals are reprinted here to facilitate comparison between these two sets.)

10.3.2.1. Gait classifier tests

10.3.2.1.1. Tests with spatio-temporal parameters

The recognition potential of the spatio-temporal parameters was studied by running the gait classifier using these parameters individually. The results are depicted in Table 10.1. The combination of all these parameters and various combinations of the parameters with the highest recognition rates were then taken. The results are shown in Table 10.2.

The following observations can be made based on these results:

- The recognition rates based on individual parameters are lower than those based on the

Table 10.1. Gait classifier recognition rates using spatio-temporal parameters

Gallery Size	Stride length	Stride time	Step length	Step time	Step width	Cadence	Walking speed	Stance phase time	Swing phase time
24	25.00%	25.00%	29.17%	62.50%	12.50%	58.33%	20.83%	29.17%	20.83%
156	2.56%	7.69%	5.77%	28.21%	2.56%	27.56%	1.92%	15.38%	8.33%

Table 10.2. Gait classifier recognition rates using various combinations of spatio-temporal parameters

Gallery Size	All spatio-temporal parameters	Step time + cadence	Step time + cadence + stance phase time
24	79.17%	58.33%	66.67%
156	47.44%	30.77%	35.90%

combination of all the parameters.

- When the two best-forming parameters (i.e., step time and cadence) are combined, the recognition rates are better than those based on individual parameters but worse than those based on the combination of all the parameters.
- When the three best-performing parameters (i.e., step time, cadence, and stance phase time) are combined, the recognition rates are better than for the above combination but worse than those based on the combination of all the parameters.
- The above observations indicate that the low-performing parameters are contributing discriminatory information which complements that contributed by the higher-performing variables.

Comparison with previous results:

The above observations are consistent with the results obtained previously with a smaller gallery of 24 subjects, except for the following: The recognition rates for the combination of the two best-performing parameters are better (rather than worse) than those based on the individual parameters.

The overall recognition rates for the larger gallery size of 156 subjects are less than those for the smaller gallery size of 24 subjects for these tests (this will continue to hold true for subsequent tests, described in the subsections following this one). However, it should be pointed out that this decrease in performance with an increase in gallery size is in accordance with what was observed in previous tests with 24 or fewer individuals.

10.3.2.1.2. Tests with angle trajectories

The recognition potential of the angle trajectory parameters was studied by running the gait classifier using these parameters individually. The results are depicted in Table 10.3. The recognition rates obtained by taking various combinations of these parameters are shown in Table 10.4.

The following observations can be made based on these results:

1) The recognition rate based on the combination of all the parameters is worse than the rates based on some of the individual parameters. This is not surprising since a combination of a large number of parameters is being taken, many of which are not yielding good results individually, and the discriminatory potential of the better-performing variables is nullified by the several low-performing ones.
2) Component 3 of the right hip angle trajectory is the best performing variable overall, yielding a recognition rate of 89.1 %.
3) When only the lower body angles are combined, the recognition rate is worse than that obtained for the combination of all parameters (87.82% v. 90.38%).

Table 10.3 (a). Gait classifier recognition rates using lower body angle trajectories

Gallery Size	Pelvis Angles			Left Foot Progress Angles			Right Foot Progress Angles			Left Hip Angles		
	1	2	3	1	2	3	1	2	3	1	2	3
24	75.00%	87.50%	70.83%	62.50%	91.67%	79.17%	41.67%	91.67%	58.33%	79.17%	91.67%	95.83%
156	37.18%	62.82%	19.23%	33.33%	63.46%	46.15%	24.36%	58.97%	29.49%	49.36%	62.18%	77.56%

Gallery Size	Right Hip Angles			Left Knee Angles			Right Knee Angles			Left Ankle Angles			Right Ankle Angles		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
24	75.00%	95.83%	100%	87.50%	95.83%	95.83%	62.50%	95.83%	91.67%	91.67%	95.83%	87.50%	79.17%	87.50%	91.67%
156	50.64%	76.28%	89.10%	63.46%	83.33%	73.08%	33.33%	73.72%	78.21%	58.97%	82.69%	62.82%	41.67%	71.79%	67.31%

Table 10.3 (b). Gait classifier recognition rates using upper body angle trajectories

Gallery Size	Thorax Angles			Neck Angles			Spine Angles		
	1	2	3	1	2	3	1	2	3
24	62.50%	50.00%	58.33%	20.83%	37.50%	25.00%	58.33%	75.00%	54.17%
156	20.51%	27.56%	26.28%	5.77%	8.33%	6.41%	32.69%	41.03%	26.28%

Gallery Size	Left Shoulder Angles			Right Shoulder Angles			Left Elbow Angles			Right Elbow Angles		
	1	2	3	1	2	3	1	2	3	1	2	3
24	62.50%	83.33%	66.67%	66.67%	79.17%	66.67%	70.83%	-	-	58.33%	-	-
156	26.92%	37.18%	40.38%	31.41%	29.49%	33.97%	36.54%	-	-	29.49%	-	-

Table 10.4. Gait classifier recognition rates using various combinations of angle trajectories (rr = individual recognition rates of angle trajectories being combined for gallery size = 24)

Gallery Size	All	Lower body	Upper body	rr>=87.5%	rr<=75%	rr<=71%	rr<=66.67%	rr<=62.5%	rr<=58.33%	rr<=54.16%	rr<=45.83%
24	79.17%	95.83%	91.67%	95.83%	91.67%	91.67%	91.67%	91.67%	87.5%	79.17%	70.83%
156	90.38%	87.82%	82.69%	94.87%	84.62%	84.62%	83.97%	80.77%	75.00%	41.67%	33.33%

4) When only the upper body angles are combined, the recognition rate is worse than that obtained for the combination of all parameters (82.69% v. 90.38%).

5) The recognition rate for the combination of all upper body angles is also worse than that for the combination of all lower body angles (82.69% v. 87.82%). This indicates that the lower body angles are more useful for recognition than the upper body angles.

6) When all the angles yielding individual recognition rates > 66.67% are combined, the resulting recognition rate is 98.7%, indicating that this might be an optimum set of variables to use for gait recognition, at least based on this data set (Note that these angles are a subset of the angles producing the best individual recognition rates (i.e., >=87.5%) for 24 subjects).

7) When the angles yielding individual recognition rates >= 87.5 % for 24 subjects are combined, the resulting recognition rate is worse than that for the combination described in 6). However, it is still better than the recognition rate obtained using the combination of all angles and only slightly less than the rate obtained for this combination using 24 subjects. This reinforces the conviction that these angles do, indeed, contribute the bulk of the discrimination potential possessed by the angle trajectories.

8) When the angles yielding individual recognition rates <=75% for 24 subjects are combined, the resulting rate is about 10 percentage points less than that obtained using the combination described in 7). As combinations of angles with progressively less individual recognition rates for 24 subjects (i.e., 66.67%, 62.5%) are considered, the resulting recognition rates fall just slightly. This indicates that even variables that yield low recognition rates individually have enough discriminatory power so that when combined, they are capable of yielding much higher recognition rates. At ~54%,

however, the recognition rate falls dramatically (by about 20%), and keeps decreasing rapidly as combinations of angles with even lower individual recognition rates are considered.

Comparison with previous results:

Observations 1), 2), 5), 7) and 8) are consistent with the results obtained previously with a smaller gallery of 24 subjects. The remaining observations differ from the results obtained using a gallery of 24 subjects as follows: When only the lower body angles or only the upper body angles are combined, the resulting recognition rate is worse (instead of better) than that obtained for the combination of all parameters.

10.3.2.1.3. Tests with combinations of spatio-temporal parameters with angle trajectories

The results in sections 10.3.2.1.1 and 10.3.2.1.2 show that the best recognition rates based on spatio-temporal parameters are obtained by using the combination of all these parameters and the best recognition rates based on combinations of angle trajectories are obtained by using a combination of the angles yielding individual recognition rates > 66.67% for the gallery of 156 subjects.

A combination of all spatio-temporal parameters and angle trajectories yielding individual recognition rates > 66.67% for the gallery of 156 subjects was, therefore, taken to explore if such a combination would yield better recognition rates. However, as shown in Table 10.5, this yields recognition rates worse than those obtained by the combination of its constituent angle trajectories alone. Also, combining all spatio-temporal parameters with all angle trajectories produces a recognition rate (shown in Table 10.5) equal to (but not better than) that using the combination of all angle trajectories alone.

Table 10.5. Gait classifier recognition rates using various combinations of spatio-temporal parameters and angle trajectories

Gallery Size	All spatio-temporal parameters + all angles with rr>66.67% for 156 subjects	All spatio-temporal parameters + all angles
24	100%	95.83%
156	92.95%	90.38%

Comparison with previous results:

The results reported above are different from the ones previously obtained for the smaller gallery of 24 subjects. For that data set, in both cases, the recognition rates produced by combining the spatio-temporal variables with angle trajectories were better than those obtained by using a combination of the angle trajectories alone.

This discrepancy in the results indicates that though the information contributed by the spatio-temporal parameters does enhance the recognition potential of the angle trajectories for smaller gallery sizes, however, as the number of subjects increases, the utility of the spatio-temporal variables for gait recognition decreases so that after the gallery size exceeds beyond a certain threshold, not only does the addition of these variables to the best-performing angle trajectories not boost the recognition rates, it actually degrades the recognition accuracy of the combination of these angles alone. Based on these observations, it can be concluded that it would be best to employ only the angle trajectories for characterizing gait when dealing with large data sets. However, the results for the gallery of 24 subjects as well as those reported in section 10.3.2.1.1 do demonstrate that the spatio-temporal parameters possess some discriminatory power. The utility of these variables for gait recognition should not, therefore, be totally ignored since these variables can still be exploited for recognizing people based on gait in scenarios where the gallery size is small or data for the angle trajectories is either not available or cannot be reliably extracted.

10.3.2.2. Integrated face-gait classifier tests

Since it is reasonable to assume that data for certain gait parameters would be available in real-world application scenarios and not for others, therefore, several different sets of gait variables were utilized by the gait classifier to evaluate the proposed system. The sets of gait variables are listed below:

1. All spatio-temporal variables
2. All angle trajectories
3. All spatio-temporal parameters and all angle trajectories
4. All angle trajectories with individual recognition rates > 66.67% for gallery size = 156
5. All angle trajectories with individual recognition rates >= 87.5% for gallery size = 24
6. All lower body angle trajectories
7. All upper body angle trajectories
8. All angle trajectories with individual recognition rates <= 75% for gallery size = 24
9. All angle trajectories with individual recognition rates <= 71% for gallery size = 24
10. All angle trajectories with individual recognition rates <= 66.67% for gallery size = 24
11. All angle trajectories with individual recognition rates <= 62.5% for gallery size = 24
12. All angle trajectories with individual recognition rates <= 58.33% for gallery size = 24
13. All angle trajectories with individual recognition rates <= 54.16% for gallery size = 24

14. All angle trajectories with individual recognition rates <= 50% for gallery size = 24

15. All angle trajectories with individual recognition rates <= 45.83% for gallery size = 24

For each set of gait variables, two integrated face-gait classifiers were constructed by combining the eigenfaces-based face classifier and the Bayesian inference-based face classifier (described in section 9.5.1 of the previous chapter), respectively, with the gait classifier. These integrated classifiers were tested using the gallery and probe sets containing all 156 subjects. For both of these integrated systems, the recognition rates using the face alone, the gait alone and the face and gait together are shown in Table 10.6. It should be noted that since both integrated systems use the same gait classifier, therefore, any difference between their performances directly follows from the difference in the performances of their face classifiers.

The following observations can be made based on these results (Note that in the following discussion, “percentage_passed” refers to the percentage of the closest matches of the face classifier that are passed to the gait classifier):

1) When the sets of gait variables whose combinations have been shown to produce the best recognition rates for gait classification (i.e., sets 2, 3, 4 and 5 in Table 10.6), are utilized, the recognition rates obtained for both integrated systems are better than those obtained using the face classifier alone. However, these rates are equal to or worse than those obtained using the gait classifier alone. This may imply that the gait classifier alone might be sufficient to recognize people using these sets of variables. However, it should be noted that these high recognition rates are being achieved using highly accurate data for these parameters. It would be unrealistic to expect such ideal rates in a real-world scenario with more noisy data and a yet larger number of subjects. Moreover, in practice,

Table 10.6. Recognition rates using face, gait and face+gait when different sets of gait variables are utilized. "Percentage passed" refers to the percentage of closest matches of the face classifier that are passed to the gait classifier

Gait variables utilized	Gait classifier	Face classifier 1	Face classifier 2	Face + gait		
				Percentage passed	Face classifier 1 + gait classifier	Face classifier 2 + gait classifier
All spatio-temporal variables	47.44%	53.21%	73.72%	10	70.51%	71.79%
				20	66.67%	63.46%
				30	63.46%	59.62%
				40	58.33%	57.69%
				50	57.69%	53.21%
				60	53.85%	51.28%
				70	51.28%	49.36%
				80	48.72%	50.00%
				90	46.79%	48.08%
All angle trajectories	90.38%	53.21%	73.72%	10	80.77%	80.13%
				20	84.62%	87.18%
				30	87.82%	87.82%
				40	88.46%	86.54%
				50	88.46%	87.82%
				60	87.82%	88.46%
				70	88.46%	88.46%
				80	89.10%	87.18%
				90	89.74%	89.10%
All spatio-temporal variables + all angle trajectories	90.38%	53.21%	73.72%	10	82.05%	81.41%
				20	84.62%	87.18%
				30	87.82%	88.46%
				40	89.74%	88.46%
				50	90.38%	89.10%
				60	88.46%	89.10%
				70	89.10%	89.10%
				80	89.10%	87.82%
				90	90.38%	89.10%
All angle trajectories with individual recognition rates > 66.67% for gallery size = 156	98.72%	53.21%	73.72%	10	86.54%	85.90%
				20	91.03%	92.95%
				30	94.87%	94.23%
				40	95.51%	96.15%
				50	97.44%	97.44%
				60	97.44%	97.44%
				70	98.08%	97.44%
				80	97.44%	98.08%
				90	98.72%	98.72%

Gait variables utilized	Gait classifier	Face classifier 1	Face classifier 2	Face + gait		
				Percentage passed	Face classifier 1 + gait classifier	Face classifier 2 + gait classifier
All angle trajectories with individual recognition rates >= 87.5% for gallery size = 24	94.87%	53.21%	73.72%	10	83.33%	83.33%
				20	86.54%	87.82%
				30	89.74%	89.74%
				40	91.67%	91.67%
				50	94.23%	94.23%
				60	94.23%	94.87%
				70	94.87%	94.87%
				80	94.23%	94.87%
				90	95.51%	94.87%
All lower body angle trajectories	87.82%	53.21%	73.72%	10	81.41%	80.77%
				20	84.62%	87.18%
				30	87.18%	89.10%
				40	88.46%	89.74%
				50	89.74%	89.10%
				60	89.10%	89.10%
				70	89.10%	89.10%
				80	87.82%	89.10%
				90	87.82%	89.10%
All upper body angle trajectories	82.69%	53.21%	73.72%	10	75.00%	75.64%
				20	79.49%	82.05%
				30	82.05%	82.05%
				40	84.62%	82.69%
				50	84.62%	83.97%
				60	83.97%	85.26%
				70	85.26%	83.33%
				80	84.62%	82.69%
				90	83.33%	82.69%
All angle trajectories with individual recognition rates <= 75% for gallery size = 24	84.62%	53.21%	73.72%	10	76.92%	76.92%
				20	80.77%	82.69%
				30	81.41%	82.69%
				40	83.33%	83.33%
				50	85.26%	85.26%
				60	84.62%	85.26%
				70	84.62%	83.97%
				80	85.26%	85.26%
				90	85.26%	85.26%

Gait variables utilized	Gait classifier	Face classifier 1	Face classifier 2	Face + gait		
				Percentage passed	Face classifier 1 + gait classifier	Face classifier 2 + gait classifier
All angle trajectories with individual recognition rates <= 71% for gallery size = 24	84.62%	53.21%	73.72%	10	76.92%	76.28%
				20	81.41%	82.05%
				30	82.05%	83.97%
				40	85.26%	83.33%
				50	85.26%	85.26%
				60	84.62%	84.62%
				70	84.62%	83.97%
				80	85.26%	84.62%
				90	85.26%	84.62%
All angle trajectories with individual recognition rates <= 66.67% for gallery size = 24	83.97%	53.21%	73.72%	10	77.56%	78.21%
				20	79.49%	82.05%
				30	82.05%	82.69%
				40	86.54%	84.62%
				50	87.18%	85.90%
				60	85.90%	87.18%
				70	86.54%	85.90%
				80	84.62%	85.90%
				90	85.90%	85.90%
All angle trajectories with individual recognition rates <= 62.5% for gallery size = 24	80.77%	53.21%	73.72%	10	78.85%	75.64%
				20	78.85%	80.13%
				30	82.69%	83.97%
				40	83.97%	82.69%
				50	84.62%	83.97%
				60	83.33%	83.97%
				70	85.26%	83.97%
				80	80.77%	80.77%
				90	81.41%	81.41%
All angle trajectories with individual recognition rates <= 58.33% for gallery size = 24	75.00%	53.21%	73.72%	10	70.51%	69.87%
				20	74.36%	74.36%
				30	78.21%	76.92%
				40	76.92%	76.92%
				50	77.56%	77.56%
				60	76.28%	76.28%
				70	75.64%	74.36%
				80	74.36%	74.36%
				90	75.00%	75.00%

Gait variables utilized	Gait classifier	Face classifier 1	Face classifier 2	Face + gait		
				Percentage passed	Face classifier 1 + gait classifier	Face classifier 2 + gait classifier
All angle trajectories with individual recognition rates <= 54.16% for gallery size = 24	41.67%	53.21%	73.72%	10	62.82%	62.18%
				20	60.90%	61.54%
				30	58.97%	55.77%
				40	55.13%	55.13%
				50	53.21%	52.56%
				60	51.28%	51.92%
				70	48.08%	47.44%
				80	44.23%	44.23%
				90	42.95%	42.31%
All angle trajectories with individual recognition rates <= 50% for gallery size = 24	45.51%	53.21%	73.72%	10	64.74%	65.38%
				20	63.46%	63.46%
				30	60.90%	58.97%
				40	58.33%	58.97%
				50	55.77%	54.49%
				60	53.85%	54.49%
				70	52.56%	50.64%
				80	48.72%	47.44%
				90	46.79%	46.15%
All angle trajectories with individual recognition rates <= 45.83% for gallery size = 24	33.33%	53.21%	73.72%	10	56.41%	58.33%
				20	51.92%	53.21%
				30	46.79%	50.00%
				40	43.59%	43.59%
				50	41.03%	41.03%
				60	39.74%	38.46%
				70	39.10%	37.18%
				80	36.54%	34.62%
				90	35.26%	33.97%

recognition would have to be performed based on the subset of variables which are available for a certain individual and it is impractical to assume that data for all these parameters would be available for all subjects in most real-world applications. So, even though we might opt to ignore the face classifier and rely exclusively on the gait classifier utilizing one of these sets of variables, however, in light of the reasons outlined above, a more pragmatic course of action would be to also examine the performance achieved by the integrated systems when sets of gait variables other than the ones mentioned above are employed.

2) When the combination of all spatio-temporal parameters is taken, both integrated systems perform better than the gait classifier alone in all cases. The first integrated system also performs better than the individual face classifier when the percentage_passed is 10 to 60%. However, the 2nd integrated classifier performs worse than the face classifier alone in all cases. This indicates that though the combination of spatio-temporal parameters aids the eigenfaces-based face classifier but it fails to enhance the performance of the more powerful Bayesian-based face classifier.

3) When the combination of all the lower body angles is taken, both integrated face-gait classifiers perform better than the individual face and gait classifiers for percentage_passed between 40 to 70%. Similarly, when the combination of all the upper body angles is taken, both integrated face-gait classifiers perform better than the individual face and gait classifiers for percentage_passed between 50 to 70%.

4) For sets 8 to 12, which are sets of lower performing variables (i.e., those with individual recognition rates <= 75%, 71%, 66.67%, 62.5%, 58.33% for gallery size = 24), both integrated face-gait classifiers perform better than the individual face and gait classifiers for certain values of percentage_passed.

5) Observation 2 also holds for sets 13 to 15, which are sets of yet lower performing variables (i.e., those with individual recognition rates <=54.16 %, 50%, 45.83% for gallery size = 24).

Comparison with previous results:

The above results differ from the results previously obtained using the gallery of 24 subjects as follows:

1. The second integrated system performs worse than the face classifier alone when the combination of all spatio-temporal variables is taken. However, this system was found to perform better than the face classifier alone when the smaller gallery of 24 subjects was utilized. This indicates that the utility of the combination of spatio-temporal parameters for recognition, used in conjunction with the Bayesian-based face classifier, declines as the number of subjects is increased. (Note, however, that the first integrated system still performs better than the individual face and gait classifiers, which is consistent with the results previously obtained).
2. Both integrated face-gait systems yield recognition rates equal to or worse than the gait classifier alone when the combination of all angle trajectories with individual recognition rates >= 87.5% for gallery size = 24 is taken. However, both these systems were found to perform better than the gait classifier alone when the smaller gallery of 24 subjects was utilized. This again indicates that the utility of the combination of these angle trajectories for recognition, used in conjunction with either face classifier, declines as the number of subjects is increased.
3. The first observation also holds for the combination of all angle trajectories with individual recognition rates <= 54.16% for gallery size = 24.

Despite these differences, it should be pointed out that the results obtained for sets 8 to 12, which are sets of lower performing variables, reflect the results previously obtained using set 13, which is a set of yet lower performing variables: both integrated face-gait systems perform better than the individual face and gait classifiers for certain values of percentage_passed. This lends credence to the conclusion reached in the previous chapter that the integrated face-gait systems yield better recognition rates than those obtained by the individual face and gait classifiers not only when some of the better-performing gait parameters are selected but also when low-performing gait variables are combined. Again, this demonstrates the real-world application potential of the proposed system and the practicability of employing such an approach even in scenarios where a limited number of gait parameters are available (and even if they are ones that do not perform well individually).

10.4. Conclusion

The system that integrates the face with the gait for automatically recognizing human beings presented in the previous chapter was evaluated using a larger data set consisting of the 24 original subjects and 132 synthetic subjects (whose gaits were generated from the 24 real subjects). A comparison of the results obtained with this larger data set with those previously obtained with the smaller data set shows that though most of the previous observations and conclusions are still true, however, the utility of certain gait variables, such as the spatio-temporal variables, for recognition does decrease as the number of subjects is increased. However, in general, the integrated face-gait system does continue to outperform the individual face and gait classifiers that it is composed of, for several different sets of gait variables, demonstrating the viability of this approach even for larger data sets.

Note that current face recognition systems have been shown to perform reasonably well when operating on high resolution face data acquired under controlled conditions [37] and our previous work [40] with the face classifiers used in this study, in particular, has confirmed this.

Moreover, our experimentation with gait recognition has revealed that if accurate data for certain sets of gait variables are available, then recognition can be performed reliably based on the gait alone. Therefore, the real challenge faced by recognition techniques employing either of these biometrics exclusively appears to lie in maintaining an adequate level of recognition accuracy when only low-quality data for these traits is available. The results of the experiments presented in this chapter and the previous one clearly demonstrate that integrating the face and the gait is a feasible solution to this problem.

CHAPTER 11

CONCLUSION

11.1. Summary

The objective of this research was to supplement and combine the face, a physical biometric, with gait, a behavioral biometric, to construct a system capable of operating effectively under a wider range of conditions than a classifier which exclusively employs only one of these biometrics. To this end, an integrated face-gait recognition system was proposed. A decision-level fusion approach was adopted where the top matches of the face classifier were passed on to the gait classifier which then determined the identity of the unknown person. For face recognition, the eigenfaces approach [6], as well as a Bayesian inference-based classifier from our previous work (described in [40]) was employed, while for gait recognition, a model-based (rather than holistic) strategy was implemented, which utilized various gait features identified as being the most pertinent for recognition based on data collected using an optoelectronic motion capture system.

This system differed from previously proposed approaches for fusing the face and the gait in the following ways: it employed a model-based (rather than silhouette-based holistic) strategy that utilizes dynamic gait features (i.e., spatio-temporal variables and joint angle trajectories), to form the gait signature. Also, it used a Bayesian inference-based face recognition technique in conjunction with the gait.

Some of the findings from the empirical evaluation of this system are outlined below:

- Both spatio-temporal variables and angle trajectories were found to be useful for recognition, though the angle trajectories yielded higher recognition rates as compared to the spatio-temporal variables.
- Step time and cadence emerged as the best-performing spatio-temporal variables while the hip angle trajectories emerged as the best-performing angle trajectories.
- The lower body angle trajectories (i.e., those of the ankles, knees and hips, foot progression and pelvis angles) were found to perform better than the upper body ones.
- When angle trajectories that produced low recognition rates individually were combined, the resulting combinations produced much higher recognition rates than the individual angles.
- The integrated face-gait classifiers outperformed the individual face and gait classifiers that they were composed of for several different sets of gait variables, thus, demonstrating the potential of using the gait to supplement the face in scenarios where the face classifier alone does not perform well due to the non-availability of high resolution face data.
- These experiments were repeated using a larger data set consisting of the 24 original subjects and 132 synthetic subjects (whose gaits were generated from the 24 real subjects). A comparison of the results obtained with this larger data set with those previously obtained with the smaller data set showed that though most of the previous observations and conclusions remained true, however, the utility of certain gait variables, such as the spatio-temporal variables, for recognition decreased as the number of subjects was increased. However, in general, the integrated face-gait system did continue to

outperform the individual face and gait classifiers that it was composed of, for several different sets of gait variables, demonstrating the viability of this approach even for larger data sets.

It should be noted that our aim, at this point, was to establish that supplementing and combining the face with the gait can produce a system capable of performing effectively in a wider range of conditions than a classifier which exclusively employs only one of these biometrics. The gait data that was used for this purpose was laboratory-based, aiming to capture the basic identity information present in various gait features. Our preliminary research has shown that integrating these biometrics is a viable option for boosting the accuracy of human recognition. If future work (examples being the suggestions listed in section 11.2) further establishes the feasibility of such an approach, the next step would be to adapt the proposed system for a real-world scenario; data acquisition and analysis will be major issues that would need to be handled in that case to ensure the successful deployment of such a system in a practical application. However, that is for later, rather than at this preliminary stage. Our implementation of the proposed system is meant as a proof of concept rather than as a practicable solution capable of being readily used in a real world application.

Automated face recognition was studied in detail during the course of this research. Experiments conducted to test the recognition accuracy of some of the most widely-used subspace-based methods (i.e., the PCA-based Eigenfaces method, the LDA-based Fisherfaces technique and the Bayesian-inference-based probabilistic Eigenfaces approach) and various combinations thereof, for different database sizes, images resolutions and number of bits per pixel, revealed the following novel results, among others:

- The recognition accuracy of the Fisherfaces method deteriorates drastically relative to the other methods with a decrease in resolution. The recognition performance of this technique is also significantly decreased when the number of bits per pixel is reduced (the other methods are relatively unaffected by the change in the number of bits per pixel).
- Certain combinations of these techniques perform better than the individual methods that those combinations are comprised of at higher resolutions.

Furthermore, a system which utilizes different resolution versions of the whole face, as well as of various facial components, in a hierarchical manner was implemented and was found to achieve higher accuracy than its single-level counterpart which uses only the highest resolution images.

11.2. Future work

We plan to continue developing this system to enhance its performance and intend to more thoroughly investigate its potential for use in real-world application scenarios. Some directions being considered for future work are outlined below:

- Due to the time-consuming nature of the motion data collection process, gait data for only a small number of subjects had been collected at the time this study was conducted. The system has, thus, currently been tested with 24 subjects. Synthetic gaits were generated by combining the real gaits to evaluate the system's performance for larger data sets as described in Chapter 10. Another option for dealing with scarcity of gait data at this point is applying the confusion metric introduced in [425] to predict the system's performance on larger populations.

- More sophisticated techniques, such as those presented in [426], may be utilized for selecting the best combination of gait features for recognition. Another option is to determine the potency of each feature for discrimination and to assign weights to the features accordingly (e.g., the F-statistic is used to rank and assign weights to various features in [420]).
- Additional gait features, such as linear and angular velocities and accelerations and correlations between angles, may be used for recognition.
- A simple nearest-neighbor classifier based on Euclidean distances is presently being used for gait classification. Other classification methods and distance measures [309] also need to be explored for recognizing the gait.
- Other strategies for fusing the face and gait, such as feature-level combination of these traits, may be investigated.
- As mentioned before, even though gait information for this study was obtained from motion capture data, our goal, ultimately, is to extract this information from video sequences. One important step in this direction is to examine how sensitive the data is to slight changes in the marker positions. For this study, the marker positions for a particular subject did not change from trial to trial. However, it is reasonable to believe that when these anatomical landmarks are located in video, the positions determined for them will vary slightly from one trial to another. Therefore, it is important to examine how gait data is affected by minor variations in the positions of these points. To this end, motion capture data for some of the subjects may be collected again. Since the marker positions will be slightly different from the ones in the previous session, they would simulate the positions extracted from video. The similarity of the gait data to that collected in the

previous session would be evaluated. Such a study would aid in assessing the applicability of the results shown here to video data.

- Since information about variables such as age, gender, limb dominance and emotional state, etc. is being collected for all gait participants, the effects of these variables on gait recognition can also be eventually studied. It should be noted that some studies have already been conducted to examine how these factors relate to gait (e.g., [427, 428] found that an individual's gender can be identified based on gait).
- Since force plate data has also been collected for all gait participants, the utility of this information for recognition can also be examined in the future. Footstep force offers the advantage of being relatively non-intrusive as compared to some other biometrics such as fingerprints and retinal scans. Moreover, unlike other biometrics, for which extracting data accurately is a major issue, footstep force information can be captured with great accuracy using force plates. The obvious disadvantage of this biometric is that it requires specialized devices for capturing it, so unlike the face and gait, which can be extracted from general surveillance video in any area, the footstep force-based recognition system can only be deployed in areas where devices for capturing this biometric have been embedded.

  More specifically, some reasons for considering the integration of the footstep force with the face for recognition include its not being dependent on subject-to-camera distance and not being directly affected by conditions that sharply degrade the performance of face recognition systems, such as large variations in illumination and facial expressions.

  Previous work on the use of footstep force as a biometric: Ground reaction force vectors using force platforms have previously been used to distinguish between healthy people and people with ankle arthrodesis [429]. These force vectors have also been measured using floor sensors (a more cost-effective solution as compared to force plates) and have been utilized for movement recognition (i.e., whether the subject is taking a step, jumping, drop-landing, sitting down, rising to stand and crouching) [430] as well as for tracking people [431].

The use of footstep force information as a biometric is still in its nascent stages. Addlesee et al. [432] were the first to utilize the vertical component of the ground reaction force vector, obtained using floor sensors, to distinguish between 15 individuals employing a HMM-based technique. They obtained 91% recognition accuracy with their system. Later, Orr et al. [433] extracted 10 features from the force profile of a single footstep to form a footstep signature and used that to classify 15 individuals and succeeded in attaining a 93% recognition rate.

Cattin et al. [434] address combining gait and footstep force information for recognition. Their system uses one feature based on the ground reaction force (i.e., the windowed Power Spectral Density of the derivative ground reaction force is used as the feature vector – note that they are extracting this from several consecutive footsteps rather than just one footstep) and three based on video data (i.e., the variance, average and two-dimensional fast fourier transform of the horizontal pixel histograms of the segmented silhouettes of one gait cycle). A new variant of generalized principal component analysis is used to project these four feature vectors into a lower-dimensional space. A method based on Bayes risk criterion is then employed to combine the four classifiers. Experiments conducted using 17 subjects yielded high recognition rates as well as low equal error rates (<0.3%) for the integrated system.

Our research on automated face recognition suggests the following directions for future work in this area:

- The utility of combining other methods based on dimensionality reduction techniques (e.g., independent components analysis [176]), employing different distance measures as similarity metrics [309], should be investigated.
- In our evaluation of the hierarchical multi-resolution face recognition system, the same algorithm was utilized for images of all resolutions. Since some preliminary experiments that we previously conducted (please see [40] for details) have shown that certain subspace-based face recognition techniques perform better than others at higher resolutions and vice versa, hence, for future research, it would be interesting to investigate if using different algorithms for different resolutions in the proposed multi-resolution approach would further enhance the recognition performance. Furthermore, the highest resolution utilized was not very large (i.e., 256x384) and only a few major facial regions were used in our experiments. It would be worthwhile to examine how using yet higher resolution images and more facial components would affect the recognition accuracy of the proposed technique.

REFERENCES

[1] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 84-90.

[2] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042-1052, 1993.

[3] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-based face recognition using mixture-distance," presented at Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1996.

[4] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," presented at Intelligent Biometric Techniques in Fingerprint and Face Recognition, 1999.

[5] Z. Liposcak and S. Loncaric, " A scale-space approach to face recognition from profiles," presented at Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns, London, UK, 1999.

[6] M. Turk and A. Pentland, "Eigenfaces For Recognition," *Journal Of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.

[7] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," presented at IEEE Conference on Computer Vision and Pattern Recognition, 1994.

[8] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.

[9] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 18, pp. 831-836, 1996.

[10] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face Recognition: A Convolutional Neural Network Approach," *IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition*, pp. 1-24, 1997.

[11] X. Lu, Y. Wang, and A. K. Jain, "Combining Classifiers for Face Recognition," presented at Proc. IEEE International Conference on Multimedia & Expo (ICME 2003), Baltimore, MD, 2003.

[12] G. L. Marcialis and F. Roli, "Fusion of appearance-based face recognition algorithms," *Pattern Analysis and Applications*, vol. 7, pp. 151-163, 2004.

[13] A. Howell and H. Buxton, "Towards unconstrained face recognition from image sequences," presented at Proceedings of the Second IEEE International Conference on Automatic Face and Gesture Recognition, 1996.

[14] T. E. de Campos, R. S. Feris, and R. M. Cesar Jr., "A Framework for Face Recognition from Video Sequences Using GWN and Eigenfeature Selection," presented at Workshop on Artificial Intelligence and Computer Vision, Atibaia, Brazil, 2000.

[15] Z. Biuk and S. Loncaric, "Face recognition from multi-pose image sequence," presented at Proceedings of 2nd IEEE R8-EURASIP Int'l Symposium on Image and Signal Processing and Analysis, Pula, Croatia, 2001.

[16] G. Gordon, "Face Recognition Based on Depth Maps and Surface Curvature," presented at SPIE Proceedings: Geometric Methods in Computer Vision, 1991.

[17] C. Beumier and M. Acheroy, "Automatic Face Recognition," in *Proceedings symposium IMAGING*. Eindhoven, The Netherlands, 2000, pp. 77-89.

[18] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland, "Illumination invariant face recognition using thermal infrared imagery," in *IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*, vol. 1. Kauai, HI, USA, 2001, pp. 527-534.

[19] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland, "Illumination invariant face recognition using thermal infrared imagery," presented at IEEE Computer Society International Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 2001.

[20] X. Chen, P. Flynn, and K. Bowyer, "IR and Visible Light Face Recognition," *Computer Vision and Image Understanding*, vol. 99, pp. 332-358, 2005.

[21] D. Zhang, Z.-H. Zhoua, and S. Chen, "Diagonal principal component analysis for face recognition," *Pattern Recognition*, vol. 39, pp. 140-142, 2006.

[22] V. D. M. Nhat and S. Lee, "Line-Based PCA and LDA Approaches for Face Recognition," in *First International Conference on Natural Computation (ICNC'05) and the Second International Conference on Fuzzy Systems and Knowledge Discovery*

(FSKD'05), vol. 3611, *Lecture Notes in Computer Science*. Changsha, China: Springer-Verlag, 2005, pp. 101-104.

[23] J. R. Price and T. F. Gee, "Face recognition using direct, weighted linear discriminant analysis and modular subspaces," *Pattern Recognition*, vol. 38, pp. 209-219, 2005.

[24] O. Toygar and A. Acan, "Multiple classifier implementation of a divide-and-conquer approach using appearance-based statistical methods for face recognition," *Pattern Recognition Letters*, vol. 25, pp. 1421-1430, 2004.

[25] Y. Yan and L. Osadciw, "Fusion in Modular Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[26] A. P. Kumar, S. Das, and V. Kamakoti, "Face Recognition Using Weighted Modular Principle Component Analysis," in *Neural Information Processing*, vol. 3316, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 362-367.

[27] W. Zhang, S. Shan, W. Gao, Y. Chang, and B. Cao, "Component-Based Cascade Linear Discriminant Analysis For Face Recognition," in *Advances in Biometric Person Authentication*, vol. 3338, *Lecture Notes in Computer Science*, 2004, pp. 288-295.

[28] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular PCA approach," *Pattern Recognition Letters*, vol. 25, pp. 429-436, 2004.

[29] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Gabor Fisher Classifier for Face Recognition," in *7th International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 91-96.

[30] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, pp. 6-21, 2003.

[31] Y. Zana, R. M. Cesar-Jr, and R. d. A. Barbosa, "Automatic Face Recognition System Based on Local Fourier-Bessel Features," in *Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 233- 240.

[32] P. J. Phillips, H. Moon, P. J. Rauss, and S. A. Rizvi, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1090-1104, 2000.

[33] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 439-446.

[34] X. Zhou and B. Bhanu, "Feature fusion of side face and gait for video-based human identification," *Pattern Recognition*, vol. 41, pp. 778-795, 2008.

[35] "http://www.montgomerycountycrimestoppers.org/nldbz.cfm." Montgomery County Crime Stoppers, Texas, U.S.A.

[36] "Plug-in Gait Product Guide - Foundation Notes (Rev 1.0)." Vicon Motion Systems, Limited, Oxford, UK.

[37] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, pp. 399-458, 2003.

[38] L. Torres, "Is there any hope for face recognition?" in *Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004)*. Lisboa, Portugal, 2004.

[39] M. S. Nixon and J. N. Carter, "Automatic Recognition by Gait," in *Proceedings of the IEEE*, vol. 94, 2006, pp. 2013-2024.

[40] R. Jafri and H. R. Arabnia, "Analysis of Subspace-based Face Recognition Techniques under Changes in Imaging Factors," in *International Conference on Information Technology (ITNG'07)*. Las Vegas, USA: IEEE Computer Society, 2007, pp. 406-413.

[41] R. Jafri and H. R. Arabnia, "Fusion of Face and Gait for Automatic Human Recognition," in *Fifth International Conference on Information Technology: New Generations (ITNG 2008)*. Las Vegas, Nevada, U.S.A.: IEEE Computer Society, 2008, pp. 167-173.

[42] C. Y. Yam, M. S. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, pp. 1057-1072, 2004.

[43] C. BenAbdelkader, R. Cutler, and L. Davis, "Person Identification using Automatic Height and Stride Estimation," in *IEEE International Conference on Pattern Recognition*, vol. 4, 2002, pp. 377 - 380.

[44] A. F. Bobick and A. Y. Johnson, "Gait recognition using static, activity-specific parameters," in *Proc. IEEE Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 423-430.

[45] R. Jafri and H. R. Arabnia, "A Survey of Component-Based Face Recognition Approaches," in *The 2007 International Conference on Artificial Intelligence (ICAI'07)*. Las Vegas, USA, 2007, pp. 103-113.

[46] P. Sinha and T. Poggio, "I think I know that face.," *Nature*, vol. 384, pp. 404, 1996.

[47] V. Goffaux and B. Rossion, "Faces are "Spatial"---Holistic Face Perception is Supported by Low Spatial Frequencies," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 32, pp. 1023-1039, 2006.

[48] J. Wang, C. Zhang, and H.-Y. Shum, "Face Image Resolution versus Face Recognition Performance Based on Two Global Methods," in *Asia Conference on Computer Vision*. Jeju Island, Korea, 2004.

[49] A. K. Jain, R. Bolle, and S. Pankanti, "Biometrics: Personal Identification in Networked Security," A. K. Jain, R. Bolle, and S. Pankanti, Eds.: Kluwer Academic Publishers, 1999.

[50] K. Kim, "Intelligent Immigration Control System by Using Passport Recognition and Face Verification," in *International Symposium on Neural Networks*. Chongqing, China, 2005, pp. 147-156.

[51] J. N. K. Liu, M. Wang, and B. Feng, "iBotGuard: an Internet-based intelligent robot security system using invariant face recognition against intruder," *IEEE Transactions On Systems Man And Cybernetics Part C-Applications And Reviews*, vol. 35, pp. 97-105, 2005.

[52] H. Moon, "Biometrics Person Authentication Using Projection-Based Face Recognition System in Verification Scenario," in *International Conference on Bioinformatics and its Applications*. Hong Kong, China, 2004, pp. 207-213.

[53] D. McCullagh, "Call It Super Bowl Face Scan 1," in *Wired Magazine*, 2001.

[54] CNN, "Education School face scanner to search for sex offenders." Phoenix, Arizona: The Associated Press, 2003.

[55] T. Choudhry, B. Clarkson, T. Jebara, and A. Pentland, "Multimodal person recognition using unconstrained audio and video," in *Proceedings, International Conference on Audio and Video-Based Person Authentication*, 1999, pp. 176-181.

[56] S. L. Wijaya, M. Savvides, and B. V. K. V. Kumar, "Illumination-tolerant face verification of low-bit-rate JPEG2000 wavelet images with advanced correlation filters for handheld devices," *Applied Optics*, vol. 44, pp. 655-665, 2005.

[57] E. Acosta, L. Torres, A. Albiol, and E. J. Delp, "An automatic face detection and recognition system for video indexing applications," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4. Orlando, Florida, 2002, pp. 3644-3647.

[58] J.-H. Lee and W.-Y. Kim, "Video Summarization and Retrieval System Using Face Recognition and MPEG-7 Descriptors," in *Image and Video Retrieval*, vol. 3115, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 179-188.

[59] C. G. Tredoux, Y. Rosenthal, L. d. Costa, and D. Nunez, "Face reconstruction using a configural, eigenface-based composite system," in *3rd Biennial Meeting of the Society for Applied Research in Memory and Cognition (SARMAC)*. Boulder, Colorado, USA, 1999.

[60] K. Balci and V. Atalay, "PCA for Gender Estimation: Which Eigenvectors Contribute?" in *Proceedings of Sixteenth International Conference on Pattern Recognition*, vol. 3. Quebec City, Canada, 2002, pp. 363-366.

[61] B. Moghaddam and M. H. Yang, "Learning Gender with Support Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 707-711, 2002.

[62] R. Brunelli and T. Poggio, "HyperBF Networks for Gender Classification," *Proceedings of DARPA Image Understanding Workshop*, pp. 311-314, 1992.

[63] A. Colmenarez, B. J. Frey, and T. S. Huang, "A probabilistic framework for embedded face and facial expression recognition," in *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition, vol. 1. Ft. Collins, CO, USA, 1999, pp. 1592-1597.

[64] Y. Shinohara and N. Otsu, "Facial Expression Recognition Using Fisher Weight Maps," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, vol. 100, 2004, pp. 499-504.

[65] F. Bourel, C. C. Chibelushi, and A. A. Low, "Robust Facial Feature Tracking," in *British Machine Vision Conference*. Bristol, 2000, pp. 232-241.

[66] K. Morik, P. Brockhausen, and T. Joachims, "Combining statistical learning with a knowledge-based approach -- A case study in intensive care monitoring," in *16th International Conference on Machine Learning (ICML-99)*. San Francisco, CA, USA: Morgan Kaufmann, 1999, pp. 268--277.

[67] S. Singh and N. Papanikolopoulos, "Vision-based detection of driver fatigue," Department of Computer Science, University of Minnesota, Technical report 1997.

[68] D. N. Metaxas, S. Venkataraman, and C. Vogler, "Image-Based Stress Recognition Using a Model-Based Dynamic Face Tracking System," *International Conference on Computational Science*, pp. 813-821, 2004.

[69] M. M. Rahman, R. Hartley, and S. Ishikawa, "A Passive And Multimodal Biometric System for Personal Identification," in *International Conference on Visualization, Imaging and Image Processing*. Spain, 2005, pp. 89-92.

[70] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 955-966, 1995.

[71] M. Viswanathan, H. S. M. Beigi, A. Tritschler, and F. Maali, "Information access using speech, speaker and face recognition," in *IEEE International Conference on Multimedia and Expo*, vol. 1, 2000, pp. 493--496.

[72] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, "Integrating Faces, Fingerprints, and Soft Biometric Traits for User Recognition," *Proceedings of Biometric Authentication Workshop, in conjunction with ECCV2004, LNCS 3087*, pp. 259-269, 2004.

[73] P. Melin and O. Castillo, "Human Recognition using Face, Fingerprint and Voice," in *Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing*, vol. 172, *Studies in Fuzziness and Soft Computing*: Springer Berlin / Heidelberg, 2005, pp. 241-256.

[74] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1160-1165, 2003.

[75] R. Chellappa, A. Roy-Chowdhury, and S. Zhou, "Human Identification Using Gait and Face," in *The Electrical Engineering Handbook*, 3rd ed: CRC Press, 2004.

[76] S. Ben-Yacoub, J. Luttin, K. Jonsson, J. Matas, and J. Kittler, "Audio-visual person verification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. Fort Collins, CO, USA, 1999, pp. 580-585.

[77] D. Bouchaffra and A. Amira, "Structural hidden Markov models for biometrics: Fusion of face and fingerprint," *Pattern Recognition*, vol. 41, pp. 852-867, 2008.

[78] H. Vajaria, T. Islam, P. Mohanty, S. Sarkar, R. Sankar, and R. Kasturi, "Evaluation and analysis of a face and voice outdoor multi-biometric system," *Pattern Recognition Letters*, vol. 28, pp. 1572-1580, 2007.

[79] Y.-F. Yao, X.-Y. Jing, and H.-S. Wong, "Face and palmprint feature level fusion for single sample biometrics recognition," *Neurocomputing*, vol. 70, pp. 1582-1586, 2007.

[80] J. Zhou, G. Su, C. Jiang, Y. Deng, and C. Li, "A face and fingerprint identity authentication system based on multi-route detection," *Neurocomputing*, vol. 70, pp. 922-931, 2007.

[81] C. Nastar and M. Mitschke, "Real time face recognition using feature combination," in *Third IEEE International Conference on Automatic Face and Gesture Recognition*. Nara, Japan, 1998, pp. 312-317.

[82] S. Gong, S. J. McKenna, and A. Psarrou., *Dynamic Vision: From Images to Face Recognition*: Imperial College Press (World Scientific Publishing Company), 2000.

[83] T. Jebara, "3D Pose Estimation and Normalization for Face Recognition," Center for Intelligent Machines, McGill University, Undergraduate Thesis May 1996 1996.

[84] P. J. Phillips, H. Wechsler, J.Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithm," *Image and Vision Computing*, vol. 16, pp. 295-306, 1998.

[85] D. Blackburn, J. Bone, and P. J. Phillips, "Face recognition vendor test 2000," Defense Advanced Research Projects Agency, Arlington, VA, Technical report A269514, Fefbruary 16, 2001 2001.

[86] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, "Face Recognition Vendor Test (FRVT 2002)," National Institute of Standards and Technology, Evaluation report IR 6965, March 2003.

[87] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, J. Czyz, L. Vandendorpe, C. McCool, S. Lowther, S. Sridharan, V.

Chandran, R. P. Palacios, E. Vidal, L. Bai, L. Shen, Y. Wang, Y.-H. Chiang, H.-C. Liu, Y.-P. Hung, A. Heinrichs, M. Müller, A. Tewes, C. v. d. Malsburg, R. P. Würtz, Z. Wang, F. Xue, Y. Ma, Q. Yang, C. Fang, X. Ding, S. Lucey, R. Goss, H. Schneiderman, N. Poh, and Y. Rodriguez, "Face Authentication Test on the BANCA Database," in *17th International Conference on Pattern Recognition*, vol. 4. Cambridge, UK, 2004, pp. 523-532.

[88] X. Q. Ding and C. Fang, "Discussions on some problems in face recognition," in *Advances In Biometric Person Authentication, Proceedings*, vol. 3338, *Lecture Notes In Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 47-56.

[89] J. Yang, X. Chen, and W. Kunz, "A PDA-based face recognition system," in *Proceedings of sixth IEEE Workshop on Applications of Computer Vision*. Orlando, Florida, 2002, pp. 19-23.

[90] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, pp. 1885-1906, 2007.

[91] M. A. Grudin, "On internal representations in face recognition systems," *Pattern Recognition*, vol. 33, pp. 1161-1177, 2000.

[92] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, pp. 6-21, 2003.

[93] T. Kanade, "Picture Processing System by Computer Complex and Recognition of Human Faces," Kyoto University, Japan, PhD. Thesis 1973.

[94] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *IEEE Computer Society Conference on Computer Vision and Templates*. San Diego, CA, USA, 1989, pp. 104-109.

[95] N. Roeder and X. Li, "Experiments in analyzing the accuracy of facial feature detection," *Vision Interface '95*, pp. 8-16, 1995.

[96] C. Colombo, A. D. Bimbo, and S. D. Magistris, "Human-computer interaction based on eye movement tracking," *Computer Architectures for Machine Perception*, pp. 258-263, 1995.

[97] M. Nixon, "Eye spacing measurement for facial recognition," in *SPIE Proceedings*, 1985, pp. 279-285.

[98] D. Reisfeld, "Generalized symmetry transforms: attentional mechanisms and face recognition," Tel-Aviv University, PhD. Thesis, technical report 1994.

[99] H. P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating faces and facial parts," in *International Workshop on Automatic Face- and Gesture-Recognition*, 1995, pp. 41-46.

[100] I. Craw, D. Tock, and A. Bennett, "Finding face features," in *Second European Conference on Computer Vision*, 1992, pp. 92-96.

[101] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-based face recognition using mixture-distance," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1996, pp. 209-216.

[102] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775-779, 1997.

[103] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. v. d. Malsburg, R. P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computers*, vol. 42, pp. 300-311, 1993.

[104] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," in *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, L. C. Jain, U. Halici, I. Hayashi, S. B. Lee, and Jae-Ho, Eds.: CRC Press, 1999, pp. 355-396.

[105] P. J. Phillips, P. Rauss, and S. Der, "FERET (FacE REcognition Technology) Recognition Algorithm Development and Test Report," U.S. Army Research Laboratory ARL-TR-995, 1996.

[106] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-recognition Algorithms," in *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 137-143.

[107] G. Sukthankar, "Face recognition: a critical look at biologically-inspired approaches," Carnegie Mellon University, Pittsburgh, PA, Technical Report: CMURI-TR-00-04 2000.

[108] P. Campadelli and R. Lanzarotti, "A Face Recognition System Based on Local Feature Characterization," in *Advanced Studies in Biometrics*, vol. 3161, *Lecture Notes in Computer Science*, M. Tistarelli, J. Bigun, and E. Grosso, Eds. Berlin: Springer, 2005, pp. 147-152.

[109] H. Shin, S. D. Kim, and H. C. Choi, "Generalized elastic graph matching for face recognition," *Pattern Recognition Letters*, vol. 28, pp. 1077–1082, 2007.

[110] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG–EBGM," *Pattern Recognition Letters*, vol. 29, pp. 1537-1543, 2008.

[111] L. D. Harmon, M. K. Khan, R. LAsch, and P. F. Raming, "Machine Identification of human faces," *Pattern Recognition*, vol. 13, pp. 97-110, 1981.

[112] L. D. Harmon, S. C. Kuo, P. F. Raming, and U. Raudkivi, "Identification of human face profiles by computers," *Pattern Recognition*, vol. 10, pp. 301-312, 1978.

[113] G. J. Kaufman and K. J. Breeding, "Automatic recognition of human faces from profile silhouettes," *IEEE Transactions On Systems Man And Cybernetics, SMC*, vol. 6, pp. 113-121, 1976.

[114] Z. Liposcak and S. Loncaric, " A scale-space approach to face recognition from profiles," in *Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns*, vol. 1689, *Lecture Notes In Computer Science*. London, UK: Springer-Verlag, 1999, pp. 243 - 250.

[115] Z. Liposcak and S. Loncaric, "Face recognition from profiles using morphological signature transform," in *Proceedings of the 21st Int'l Conference Information Technology Interfaces*. Pula, Croatia, 1999, pp. 93-98.

[116] R. Brunelli and T. Poggio, "Face Recognition Through Geometrical Features," in *Proceedings of the Second European Conference on Computer Vision*, vol. 588, *Lecture Notes In Computer Science*, G. Sandini, Ed. London, UK: Springer-Verlag, 1992, pp. 782-800.

[117] R. Cendrillon and B. C. Lowell, "Real-Time Face Recognition using Eigenfaces," in *Proceedings of the SPIE International Conference on Visual Communications and Image Processing*, vol. 4067, 2000, pp. 269-276.

[118] R. J. Baron, "Mechanisms of Human Facial Recognition," *International Journal of Man-Machine Studies*, vol. 15, pp. 137-178, 1981.

[119] R.-J. J. Huang, "Detection Strategies for face recognition using learning and evolution," George Mason University, Fairfax, Virginia, Ph. D. Dissertation 1998.

[120] L. Sirovich and M. Kirby, "Low-dimensional Procedure for the Characterization of Human Faces," *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 4, pp. 519-524, 1987.

[121] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. New Jersey: Prentice-Hall, 1988.

[122] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second ed. Boston, MA: Academic Press, 1990.

[123] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586-591.

[124] Y. Moses, Y. Adini, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," in *European Conf. Computer Vision*, 1994, pp. 286-296.

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

[126] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 228-233, 2001.

[127] B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian Similarity Measure for Direct Image Matching," in *Proceedings 13th International Conference on Pattern Recognition*, 1996, pp. 350-358.

[128] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 19, pp. 696-710, 1997.

[129] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear Subspace Analysis of Image Ensembles," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2003, pp. 93-99.

[130] Q. Yang and X. Q. Ding, "Symmetrical Principal Component Analysis and Its Application in Face Recognition," *Chinese Journal of Computers*, vol. 26, pp. 1146–1151, 2003.

[131] J. Yang and D. Zhang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, pp. 131-137, 2004.

[132] J. Meng and W. Zhang, "Volume measure in 2DPCA-based face recognition," *Pattern Recognition Letters*, vol. 28, pp. 1203-1208, 2007.

[133] G. D. C. Cavalcanti and E. C. B. C. Filho, "Eigenbands Fusion for Frontal Face Recognition," in *Proceedings of IEEE Internationall Conference on Image Processing*, vol. 1, 2003, pp. 665–668.

[134] K. R. Tan and S. C. Chen, "Adaptively weighted sub-pattern PCA for face recognition," *Neurocomputing*, vol. 64, pp. 505-511, 2005.

[135] V. D. M. Nhat and S. Lee, "An Improvement on PCA Algorithm for Face Recognition," in *Advances in Neural Networks - ISNN 2005*, vol. 3498, *Lecture Notes in Computer Science*. Chongqing: Springer, 2005, pp. 1016-1021.

[136] N. Sun, H.-x. Wang, Z.-h. Ji, C.-r. Zou, and L. Zhao, "An efficient algorithm for Kernel two-dimensional principal component analysis," *Neural Computing & Applications*, vol. 17, pp. 59-64, 2008.

[137] 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.

[138] F. Song, D. Zhang, J. Wang, H. Liu, and Q. Tao, "A parameterized direct LDA and its application to face recognition," *Neurocomputing*, vol. 71, pp. 191-196, 2007.

[139] D. Zhou and X. Yang, "Face Recognition Using Direct-Weighted LDA," in *8th Pacific Rim International Conference on Artificial Intelligence*. Auckland, New Zealand, 2004, pp. 760-768.

[140] L. Chen, H. Liao, M. Ko, L. J., and G. Yu, " A New LDA-based Face Recognition System Which Can Solve the Small Samples Size Problem," *Journal of Pattern Recognition*, vol. 33, pp. 1713–1726, 2000.

[141] W. Liu, Y. Wang, S. Z. Li, and T. Tan, "Null Space Approach of Fisher Discriminant Analysis for Face Recognition," in *Biometric Authentication*, vol. 3087, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 32-44.

[142] X. Wang and X. Tang, "Dual-space Linear Discriminant Analysis for Face Recognition," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2004, pp. 564–569.

[143] M. Loog, R. P. W. Duin, and R. Haeb-Umbach, "Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 762-766, 2001.

[144] J. H. Friedman, "Regularized Discriminant Analysis," *Journal of the American Statistical Association*, vol. 84, pp. 165-175, 1989.

[145] P. Howland and H. Park, "Generalized Discriminant Analysis Using the Generalized Singular Value Decomposition," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 26, pp. 995–1006, 2004.

[146] J. P. Ye, R. Janardan, C. H. Park, and H. Park, "An Optimization Criterion for Generalized Discriminant Analysis on Undersampled Problems," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 26, pp. 982–994, 2004.

[147] J. W. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face Recognition Using LDA-based Algorithms," *IEEE Trans. On Neural Networks*, vol. 14, pp. 195–200, 2003.

[148] J. W. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Boosting Linear Discriminant Analysis for Face Recognition," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, 2003, pp. 657–660.

[149] Q. Yang and X. Q. Ding, "Discriminant Local Feature Analysis of Facial Images," in *IEEE International Conference on Image Processing*, vol. 2, 2003, pp. 863–866.

[150] Q. Liu, H. Lu, and S. Ma, "Improving Kernel Fisher Discriminant Analysis for Face Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 42-49, 2004.

[151] B. Schölkopf, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.

[152] Q. Liu, X. Tang, H. Lu, and S. Ma, "Kernel Scatter-Difference Based Discriminant Analysis for Face Recognition," in *Proc. IEEE International Conference on Pattern Recognition*, 2004, pp. 419–422.

[153] M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recognition Letters*, vol. 26, pp. 527-532, 2005.

[154] H. L. Xiong, M. N. S. Swamy, and M. O. Ahmad, "Two-dimensional FLD for face recognition," *Pattern Recognition*, vol. 38, pp. 1121-1124, 2005.

[155] X. Y. Jing, Y. Y. Tang, and D. Zhang, "A Fourier-LDA approach for image recognition," *Pattern Recognition*, vol. 38, pp. 453-457, 2005.

[156] Y. W. Pang, L. Zhang, M. J. Li, Z. K. Liu, and W. Y. Ma, "A novel Gabor-LDA based face recognition method," in *Advances In Multimedia Information Processing - Pcm 2004, Pt 1, Proceedings*, vol. 3331, *Lecture Notes In Computer Science*, 2004, pp. 352-358.

[157] V. D. M. Nhat and S. Lee, "Block LDA for Face Recognition," in *Computational Intelligence and Bioinspired Systems*, vol. 3512, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 899-905.

[158] D. Zhou and X. Yang, "Face Recognition Using Enhanced Fisher Linear Discriminant Model with Facial Combined Feature," in *PRICAI 2004: Trends in Artificial Intelligence*, vol. 3157, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 769-777.

[159] W. C. Zhang, S. G. Shan, W. Gao, Y. Z. Chang, and B. Cao, "Component-based cascade linear discriminant analysis for face recognition," in *Advances In Biometric Person Authentication, Proceedings*, vol. 3338, *Lecture Notes In Computer Science*, 2004, pp. 288-295.

[160] H. Zhao and P. C. Yuen, "Incremental Linear Discriminant Analysis for Face Recognition," *IEEE Transactions on Systems, Man & Cybernetics: Part B*, vol. 38, pp. 210-221, 2008.

[161] Y. Chang, C. Hu, and M. Turk, "Manifold of facial expression," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003, pp. 28-35.

[162] K.-C. Lee, J. Ho, M.-H. Yang, and D. J. Kriegman, "Video-Based Face Recognition Using Probabilistic Appearance Manifolds," *Computer Vision and Pattern Recognition*, vol. 1, pp. 313-320, 2003.

[163] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[164] S. Roweis, L. Saul, and G. E. Hinton, "Global coordination of local linear models," *Advances in Neural Information Processing Systems*, vol. 14, pp. 889-896, 2002.

[165] H. Seung and D. Lee, "The Manifold Ways of Perception," *Science*, vol. 290, pp. 2268-2269, 2000.

[166] A. Shashua, A. Levin, and S. Avidan, "Manifold Pursuit: A New Approach to Appearance Based Recognition," in *International Conference on Pattern Recognition*, vol. 3, 2002, pp. 590-594.

[167] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[168] L. K. Saul and S. T. Roweis, "Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds," *Journal of Machine Learning Research*, vol. 4, pp. 119-155, 2003.

[169] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, pp. 585-591, 2001.

[170] X. He, S. C. Yan, Y. X. Hu, and H. J. Zhang, "Learning a Locality Preserving Subspace for Visual Recognition," in *Proceedings of 9th IEEE International Conference on Computer Vision*, vol. 1, 2003, pp. 385- 392.

[171] S. C. Yan, H. J. Zhang, Y. X. Hu, B. Y. Zhang, and Q. S. Cheng, "Discriminant Analysis on Embedded Manifold," in *European Conference on Computer Vision*, vol. LNCS 3021: Springer Berlin / Heidelberg, 2004, pp. 121–132.

[172] J. Zhang, S. Z. Li, and J. Wang, "Nearest Manifold Approach for Face Recognition," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 223–228.

[173] Y. Wu, K. L. Chan, and L. Wang, "Face Recognition based on Discriminative Manifold Learning," in *Proc. IEEE Int'l Conf. on Pattern Recognition*, vol. 4, 2004, pp. 171–174.

[174] X. He, S.-C. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 328-340, 2005.

[175] P. Comon, "Independent component analysis—A new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.

[176] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1450-1464, 2002.

[177] B. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, "Recognizing faces with PCA and ICA," *Computer Vision and Image Understanding: Special Issue on Face Recognition*, vol. 91, pp. 115-137, 2003.

[178] C. Liu and H. Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition," in *International Conference on Audio and Video Based Biometric Person Authentication*. Washington, D.C., 1999, pp. 211-216.

[179] J. Kim, J. Choi, and J. Yi, "Face Recognition Based on Locally Salient ICA Information," in *Biometric Authentication Workshop*, vol. 3087, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 1-9.

[180] J. Yi, J. Kim, J. Choi, J. Han, and E. Lee, "Face Recognition Based on ICA Combined with FLD," *Biometric Authentication*, pp. 10-18, 2002.

[181] T. Martiriggiano, M. Leo, T. D'Orazio, and A. Distante, "Face Recognition by Kernel Independent Component Analysis," in *The 18th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*. Bari, Italy, 2005, pp. 55-58.

[182] J. Kim, J. Choi, and J. Yi, "ICA Based Face Recognition Robust to Partial Occlusions and Local Distortions," in *International Conference on Bioinformatics and its Applications*. Fort Lauderdale, Florida, USA, 2004, pp. 147-154.

[183] K.-C. Kwak and W. Pedrycz, "Face Recognition Using an Enhanced Independent Component Analysis Approach," *IEEE Transactions on Neural Networks*, vol. 18, pp. 530-541, 2007.

[184] N. H. Foon, A. T. B. Jin, and D. N. C. Ling, "Face recognition using wavelet transform and non-negative matrix factorization," in *Ai 2004: Advances In Artificial Intelligence, Proceedings*, vol. 3339, *Lecture Notes In Artificial Intelligence*, 2004, pp. 192-202.

[185] D. D. Lee and H. S. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature*, vol. 401, pp. 788-791, 1999.

[186] W. Liu, Y. Wang, S. Z. Li, and T. Tan, "Nearest Intra-Class Space Classifier for Face Recognition," in *The 17th International Conference on Pattern Recognition (ICPR)*, vol. 4. Cambridge, UK, 2004, pp. 495-498.

[187] J. Li, S. Zhou, and C. Shekhar, "A Comparison of Subspace Analysis for Face Recognition," in *Proc. IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing*, 2003, pp. 121–124.

[188] Q. Yang and X. Tang, "Recent Advances in Subspace Analysis for Face Recognition," *SINOBIOMETRICS*, pp. 275-287, 2004.

[189] D. DeMers and G. W. Cottrell, "Non-linear dimensionality reduction," *Advances in Neural Information Processing Systems*, vol. 5, pp. 580-587, 1993.

[190] J. Weng, N. Ahuja, and T. S. Huang, "Learning recognition and segmentation of 3-D objects from 3-D images," in *Proceedings of the International Conference on Computer Vision (ICCV 93)*. Berlin, Germany, 1993, pp. 121-128.

[191] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, pp. 1464-1480, 1990.

[192] T. Kohonen, *Self-organizing maps*. Berlin, Germany: Springer-Verlag, 1995.

[193] A. Eleyan and H. Demirel, "Face Recognition System Based on PCA and Feedforward Neural Networks," in *Computational Intelligence and Bioinspired Systems*, vol. 3512, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 935-942.

[194] B. Li and H. Yin, "Face Recognition Using RBF Neural Networks and Wavelet Transform," in *Advances in Neural Networks – ISNN 2005*, vol. 3497, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 105-111.

[195] P. Melin, C. Felix, and O. Castillo, "Face recognition using modular neural networks and the fuzzy Sugeno integral for response integration," *International Journal Of Intelligent Systems*, vol. 20, pp. 275-291, 2005.

[196] G. C. Zhang, X. S. Huang, S. Z. Li, Y. S. Wang, and X. H. Wu, "Boosting local binary pattern (LBP)-based face recognition," in *Advances In Biometric Person Authentication, Proceedings*, vol. 3338, *Lecture Notes In Computer Science*, 2004, pp. 179-186.

[197] T. Ojala, M. Pietikainen, and M. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification width local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

[198] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119-139, 1997.

[199] T. Ahonen, A. Hadid, and M.Pietikainen, "Face recognition with local binary patterns," in *Proceedings of the European Conference on Computer Vision*, vol. 3021, *Lecture Notes in Computer Science*. Prague, Czech Republic: Springer, 2004, pp. 469–481.

[200] U. Krebel, "Pairwise classification and support vector machines," *Advance in Kernel Methods – Support Vector Learning*, pp. 255-268, 1999.

[201] T. Hastie and R. Tibshirani, "Classification by Pairwise Coupling," *The Annals of Statistics*, vol. 26, pp. 451-471, 1998.

[202] M. Moreira and E. Mayoraz, "Improved Pairwise Coupling Classification with Correcting Classifiers," in *Proceedings of the 10th European Conference on Machine Learning*, vol. 1398, *Lecture Notes In Computer Science*. London, UK: Springer-Verlag, 1998, pp. 160-171.

[203] H. Li, F. Qi, and S. Wang, "Face Recognition with Improved Pairwise Coupling Support Vector Machines," in *Computational Intelligence and Bioinspired Systems*, vol. 3512, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 927-934.

[204] Z. Li and S. Tang, "Face Recognition Using Improved Pairwise Coupling Support Vector Machines," in *Proc. of Intl. Conf. on Neural Information Processing*, vol. 2, 2002, pp. 876-880.

[205] J. Platt, "Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods," in *Advances in Large Margin Classifiers*, A. J. Smola, P. L. Bartlett, B. Sch¨olkopf, and D. Schuurmans, Eds.: MIT Press, 2000, pp. 61-74.

[206] H. Q. Li, S. Y. Wang, and F. H. Qi, "Automatic face recognition by support vector machines," in *Combinatorial Image Analysis, Proceedings*, vol. 3322, *Lecture Notes In Computer Science*, 2004, pp. 716-725.

[207] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–267, 1998.

[208] G. Dai and C. Zhou, "Face Recognition Using Support Vector Machines with the Robust Feature," in *Proceedings of IEEE Workshop on Robot and Human Interactive Communication*, 2003, pp. 49-53.

[209] O. D´eniz, M. Castrill´on, and M. Hern´andez, "Face Recognition Using Independent Component Analysis and Support Vector Machines," *Pattern Recognition Letters*, vol. 24, pp. 2153-2157, 2003.

[210] K. Jonsson, J. Matas, J. Kittler, and Y. P. Li, "Learning Support Vector Machines for Face Verification and Recognition," in *Proceedings of IEEE International Conference on Automatic and Gesture Recognition*, 2000, pp. 208-213.

[211] G. Guo, S. Li, and C. Kapluk, "Face Recognition by Support Vector Machines," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA, 2000, pp. 196-201.

[212] Y. Liang, W. Gong, Y. Pan, W. Li, and Z. Hu, "Gabor Features-Based Classification Using SVM for Face Recognition," in *Advances in Neural Networks – ISNN 2005*, vol. 3497, *Lecture Notes in Computer Science*. Chongqing: Springer, 2005, pp. 118-123.

[213] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," in *Readings in Speech Recognition*, vol. 77, A. Waibel and K. Lee, Eds. San Francisco, CA: Morgan Kaufmann, 1989, pp. 257-285.

[214] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*. Sarasota, FL, USA, 1994, pp. 138-142.

[215] F. S. Samaria, "Face recognition using Hidden Markov Models," Trinity College, University of Cambridge, Cambridge, UK, Ph. D. Thesis 1994.

[216] A. V. Nefian and M. H. Hayes III, "Face Recognition using an embedded HMM," in *IEEE International Conference Audio Video Biometric based Person Authentication*, 1999, pp. 19-24.

[217] S. Kuo and O. Agazzi, "Keyword Spotting in poorly printed documents using pseudo 2-D Hidden Markov Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 842-848, 1994.

[218] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 570-582, 2000.

[219] H.-L. Huang, H.-M. Chen, S.-J. Ho, and S.-Y. Ho, "Advanced Evolutionary Pursuit for Face Recognition," *accepted by Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology*, 2006.

[220] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based Discriminant Learning with Boosting for Face Recognition," *IEEE Transactions on Neural Networks*, vol. 17, pp. 166-178, 2006.

[221] J. Lu and K. N. Plataniotis, "Boosting face recognition on a large-scale database," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2. Rochester, NY, 2002, pp. 109-112.

[222] R. E. Schapire, "The boosting approach to machine learning: An overview," in *MSRI Workshop Nonlinear Estimation and Classification*, 2002, pp. 149–172.

[223] F. Roli and J. Kittler, "Multiple Classifier Systems, Third International Workshop, MCS 2002, Cagliari, Italy, June 24-26, 2002, Proceedings," in *Lecture Notes in Computer Science*, vol. 2364, *Lecture Notes in Computer Science*: Springer Verlag, 2002.

[224] X. Lu, Y. Wang, and A. K. Jain, "Combining Classifiers for Face Recognition," in *Proc. IEEE International Conference on Multimedia & Expo (ICME 2003)*. Baltimore, MD, 2003, pp. 13-16.

[225] C. M. Bishop, *Neural Networks for Pattern Recognition*: Oxford University Press, UK, 1995.

[226] G. L. Marcialis and F. Roli, "Fusion of LDA and PCA for face recognition," in *Proceedings of the Workshop on Machine Vision and Perception, 8th Workshop of the Italian Association for Artificial Intelligence (ALLA 02)*, 2002.

[227] G. L. Marcialis and F. Roli, "Fusion of LDA and PCA for face verification," in *Proceedings of the Workshop on Biometric Authentication*, vol. 2359, *LNCS*, M. Tistarelli, J. Bigun, and A. K. Jain, Eds. Copenhagen, Denmark: Springer-Verlag, 2002, pp. 30-37.

[228] B. Achermann and H. Bunke, "Combination of Classifiers on the Decision Level for Face Recognition," Insitut fur Informatik und angewandte Mathematik, Universitat Bern, Bern, Germany, Technical Report IAM-96-002 January 1996 1996.

[229] A. S. Tolba and A. N. Abu-Rezq, "Combined Classifier for Invariant Face Recognition," *Pattern Analysis and Applications*, vol. 3, pp. 289-302, 2000.

[230] Y. H. Wan, S. M. Ji, Y. Xie, X. Zhang, and P. J. Xie, "Video program clustering indexing based on face recognition hybrid model of hidden Markov model and support vector machine," in *Combinatorial Image Analysis, Proceedings*, vol. 3322, *Lecture Notes In Computer Science*, 2004, pp. 739-749.

[231] K. C. Kwak and W. Pedrycz, "Face recognition: A study in information fusion using fuzzy integral," *Pattern Recognition Letters*, vol. 26, pp. 719-733, 2005.

[232] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets System*, vol. 29, pp. 201-227, 1988.

[233] J. Haddadnia, K. Faez, and M. Ahmadi, "N-Feature Neural Network Human Face Recognition," *Image and Vision Computing*, vol. 22, pp. 1071-1082, 2002.

[234] J. Haddadnia, K. Faez, and P. Moallem, "Neural network based face recognition with moment invariants," in *IEEE International Conference on Image Processing*, vol. 1. Thessaloniki, Greece, 2001, pp. 1018-1021.

[235] J. Haddadnia, M. Ahmadi, and K. Faez, "An Efficient Method for Recognition of Human Face Recognition Using Higher Order Pseudo Zernike Moment Invariant," in *The 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition*. Washington, DC, USA, 2002.

[236] L. Torres, L. Lorente, and J. Vilà, "Face recognition using self-eigenfaces," in *International Symposium on Image/Video Communications Over Fixed and Mobile Networks*. Rabat, Morocco, 2000, pp. 44-47.

[237] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, pp. 705-740, 1995.

[238] A. Howell and H. Buxton, "Towards unconstrained face recognition from image sequences," in *Proceedings of the Second IEEE International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 224-229.

[239] J. Moody and C. Darken, "Learning with localized receptive fields," in *Proceedings of the 1988 Connectionist Models Summer School*, D. Touretzky, G. Hinton, and T. Sejnowski, Eds.: Morgan Kaufmann, 1988, pp. 133-143.

[240] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, pp. 281-294, 1989.

[241] S. McKenna and S. Gong, "Combined motion and model-based face tracking," in *Proceedings of British Machine Vision Conference*. Edinburgh, UK, 1996, pp. 755-765.

[242] T. E. de Campos, R. S. Feris, and R. M. Cesar Jr., "A Framework for Face Recognition from Video Sequences Using GWN and Eigenfeature Selection," in *Workshop on Artificial Intelligence and Computer Vision*. Atibaia, Brazil, 2000.

[243] R. S. Feris, T. E. Campos, and R. M. Cesar Jr., "Detection and Tracking of Facial Features in Video Sequences," in *Mexican International Conference on Artificial Intelligence (MICAI 2000)*, vol. 1793, *Lecture Notes in Artificial Intelligence*. Acapulco, Mexico: Springer-Verlag, 2000, pp. 129-137.

[244] V. Kruger and G. Sommer, "Affine real-time face tracking using a wavelet network," in *ICCV'99 Workshop: Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*. Corfu, Greece, 1999, pp. 141-148.

[245] T. E. de Campos, I. Bloch, and R. M. Cesar Jr., "Feature Selection Based on Fuzzy Distances Between Clusters: First Results on Simulated Data," in *Proceedings of International Conference on Advances on Pattern Recognition-ICAPR'2000*, vol. 2013, *Lecture Notes In Computer Science*. Rio de Janeiro, Brasil: Springer-Verlag, 2001, pp. 186-195.

[246] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York, USA: Wiley, 1973.

[247] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4-37, 2000.

[248] Z. Biuk and S. Loncaric, "Face recognition from multi-pose image sequence," in *Proceedings of 2nd IEEE R8-EURASIP Int'l Symposium on Image and Signal Processing and Analysis*. Pula, Croatia, 2001, pp. 319-324.

[249] V. Krueger and S. Zhou, "Exemplar-based face recognition from video," in *Computer Vision - ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002. Proceedings, Part IV*, vol. 2353, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2002, pp. 732.

[250] S. Zhou, V. Krueger, and R. Chellappa, "Face Recognition from Video: A CONDENSATION Approach," in *Proc. of Fifth 1EEE International Conference on Automatic Face and Gesture Recognition*. Washington D.C., USA, 2002, pp. 221-228.

[251] G. L. Marcialis and F. Roli, "Fusion of face recognition algorithms for video-based surveillance systems," in *Multisensor Surveillance Systems: The Fusion Perspective*, G. L. Foresti, C. S. Regazzoni, and P. K. Varshney, Eds.: Kluwer, 2003, pp. 235-250.

[252] J. Steffens, E. Elagin, and H. Neven, "Person Spotter – fast and robust system for human detection, tracking and recognition," in *Proceedings, International Conference on Audio- and Video-Based Person Authentication*, 1999, pp. 96-101.

[253] presented at Proceedings of the International Conferences on Audio- and Video-Based Person Authentication, 1997-2005.

[254] S. Zhou and R. Chellappa, "Beyond a single still image: Face recognition from multiple still images and videos," in *Face Processing: Advanced Modeling and Methods*: Academic Press, 2005.

[255] M. C. Chiang and T. E. Boult, "Local Blur Estimation and Super-resolution," in *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 821-826.

[256] K. Aizawa, T. Komatsu, and T. Saito, "A scheme for acquiring very high resolution images using multiple cameras," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, 1992, pp. 289-292.

[257] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646-1658, 1997.

[258] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia, "Reconstruction of high resolution 3D visual information using sub-pixel camera displacements," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 654-657.

[259] M. H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, 2002.

[260] A. Tibbalds, "Three Dimensional Human Face Acquisition for Recognition," Trinity College, University of Cambridge, Cambridge, Ph. D. Thesis March 1998.

[261] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Proceedings of the 7th IEEE International Symposium on Signal Processing and Its Applications*, vol. 2, 2003, pp. 201-204.

[262] G. Gordon, "Face Recognition Based on Depth Maps and Surface Curvature," in *SPIE Proceedings: Geometric Methods in Computer Vision*, vol. 1570, 1991, pp. 234-- 247.

[263] Cyberware, "Cyberware Inc.: Electronic Documentation."

[264] Scanners, "3D Scanners Ltd. Electronic Documentation, from http://www.3dscanners.com/.HTML."

[265] U. Castellani, M. Bicego, G. Iacono, and V. Murino, "3D Face Recognition Using Stereoscopic Vision," in *Advanced Studies in Biometrics*, vol. 3161, *Lecture Notes in Computer Science*, M. Tistarelli, J. Bigun, and E. Grosso, Eds.: Springer Berlin / Heidelberg, 2005, pp. 126-137.

[266] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel B-Splines," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, pp. 228-244, 1997.

[267] G. Pan, Z. Wu, and Y. Pan., "Automatic 3D face verification from range data," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 2003, pp. 193-196.

[268] C. Xu, Y. Wang, T. Tan, and L. Quan, "Automatic 3D face recognition combining global geometric features with local shape variation information," in *International Conference on Automated Face and Gesture Recognition*, 2004, pp. 308–313.

[269] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn, "Local feature based 3D face recognition," in *Audio- and Video-Based Biometric Person Authentication*, vol. 3546, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 909-918.

[270] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Integration of local and global geometrical cues for 3D face recognition," *Pattern Recognition*, vol. 41, pp. 1030-1040, 2008.

[271] J. Y. Cartoux, J. T. LaPreste, and M. Richetin, "Face authentication or recognition by profile extraction from range images," in *Proceedings of the Workshop on Interpretation of 3D Scenes*, 1989, pp. 194-199.

[272] T. Nagamine, T. Uemura, and I. Masuda, "3D facial image analysis for human identification," in *Proceedings of International Conference on Pattern Recognition (ICPR' 1992)*. The Hague, Netherlands, 1992, pp. 324-327.

[273] C. Li and A. Barreto, "Profile-Based 3D Face Registration and Recognition," in *Advances in Neural Networks – ISNN 2005*, vol. 3497, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 478-488.

[274] Y. Wu, G. Pan, and Z. Wu, "Face Authentication based on Multiple Profiles Extracted from Range Data," in *4th International Conference on Audio- and Video-based Biometric Person Authentication*, vol. LNCS-2688, 2003, pp. 515-522.

[275] B. Gokberk, A. A. Salah, and L. Akarun, "Rank-Based Decision Fusion for 3D Shape-Based Face Recognition," in *Proceedings of the 5th International Conference on Audio and Video-based Biometric Person Authentication*, vol. 3546, *Lecture Notes in Computer Science*, 2005, pp. 1019-1028.

[276] J.-G. Wang, K.-A. Toh, and R. Venkateswarlu, "Fusion of Appearance and Depth Information for Face Recognition," in *Fifth International Conference on Audio- and Video-Based Biometric Person Authentication*, vol. 3546, *Lecture Notes in Computer Science*: Springer, 2005, pp. 919-928.

[277] F. Tsalakanidou, D. Tzocaras, and M. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 24, pp. 1427–1435, 2003.

[278] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *Multimodal User Authentication Workshop*. Nice, France, 2003, pp. 25-32.

[279] C. Beumier and M. Acheroy, "Face verification from 3D and grey level clues," *Pattern Recognition Letters*, vol. 22, pp. 1321-1329, 2001.

[280] P. Besl and N. McKay, "A method for registration of 3D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 239–256, 1992.

[281] T. Papatheodorou and D. Reuckert, "Evaluation of automatic 4D face recognition using surface and texture registration," in *International Conference on Automated Face and Gesture Recognition*, 2004, pp. 321–326.

[282] X. Lu and A. K. Jain, "Integrating range and texture information for 3D face recognition," in *Proceedings of 7th IEEE Workshop on Applications of Computer Vision*. Breckenridge, CO, 2005, pp. 156-163.

[283] Y. Wang, C. Chua, and Y. Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recognition Letters*, vol. 23, pp. 1191-1202, 2002.

[284] C. Benabdelkader and P. A. Griffin, "Comparing and combining depth and texture cues for face recognition," *Image And Vision Computing*, vol. 23, pp. 339-352, 2005.

[285] A. Ruifrok, A. Scheenstra, and R. C. Veltkamp, "A Survey of 3D Face Recognition Methods," in *Audio- and Video-based Biometric Person Authentication*, vol. 3546, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2005, pp. 891-899.

[286] K. W. Bowyer, K. Chang, and P. J. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, pp. 1-15, 2006.

[287] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland, "Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery," in *CVPR workshop on Computer Vision Beyond Visual Spectrum*. Kauai, HI, USA, 2001.

[288] R. Cutler, "Face recognition using infrared images and eigenfaces," University of Maryland at College Park, College Park, MD, USA, Technical report CSC 989, 1996.

[289] A. Gyaourova, G. Bebis, and I. Pavlidis, "Fusion of Infrared and Visible Images for Face Recognition," in *Computer Vision - ECCV 2004*, vol. 3024, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 456-468.

[290] A. Selinger and D. A. Socolinsky, "Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study," Equinox Corporation, Technical Report 02-01 February 2002.

[291] S.-Q. Wu, W. Song, L.-J. Jiang, S. L. Xie, F. Pan, W.-Y. Yau, and S. Ranganath, "Infrared Face Recognition by Using Blood Perfusion Data," in *Audio- and Video-Based Biometric Person Authentication,5th International Conference, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005, Proceedings*, vol. 3546, *Lecture Notes in Computer Science*: Springer, 2005, pp. 320-328.

[292] J. Wilder, P. J. Phillips, C. H. Jiang, and S. Wiener, "Comparison of Visible and Infra-red Imagery for Face Recognition," in *Proceedings, International Conference on Automatic Face and Gesture Recognition*. Killington, VT, USA, 1996, pp. 182-187.

[293] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja, "Memory-based Face Recognition for Visitor Identification," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 214-220.

[294] X. Chen, P. Flynn, and K. Bowyer, "Visible-light and infrared face recognition," in *Proceedings of the Workshop on Multimodal User Authentication*. Santa Barbara, California, USA, 2003, pp. 48–55.

[295] Identix, "Identix Inc.: Electronic Documentation."

[296] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226-239, 1998.

[297] R. Singh, M. Vatsa, and A. Noore, "Hierarchical fusion of multi-spectral face images for improved recognition performance," *Information Fusion*, vol. 9, pp. 200-210, 2008.

[298] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *Proceedings of the IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in Conjunction with CVPR, 2004*, 2004, pp. 94–99.

[299] R. Singh, M. Vatsa, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition," *Pattern Recognition*, vol. 41, pp. 880-893, 2008.

[300] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition - a review," *Computer Vision And Image Understanding*, vol. 97, pp. 103-135, 2005.

[301] Q. Yang and X. O. Tang, "Recent advances in subspace analysis for face recognition," in *Advances In Biometric Person Authentication, Proceedings*, vol. 3338, *Lecture Notes In Computer Science*, 2004, pp. 275-287.

[302] B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian Similarity Measure for Direct Image Matching," *Proceedings 13th International Conference on Pattern Recognition*, pp. 350-358, 1996.

[303] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, pp. 336-341, 1998.

[304] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.

[305] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1090-1104, 2000.

[306] R. Gross, S. Baker, I. Matthews, and T. Kanade, "Face Recognition Across Pose and Illumination," in *Handbook of Face Recognition*, S. Z. L. a. A. K. Jain, Ed., June 2004.

[307] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.

[308] P. Navarrete and J. Ruiz-Del-Solar, "Analysis and comparison of eigenspace-based face recognition approaches," *International Journal Of Pattern Recognition And Artificial Intelligence*, vol. 16, pp. 817-830, 2002.

[309] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recognition Letters*, vol. 25, pp. 711-724, 2004.

[310] "FERET Database http://www.itl.nist.gov/iad/humanid/feret/," NIST, 2001.

[311] G. H. Givens, J. R. Beveridge, B. A. Draper, and D. Bolme, "A Statistical Assessment of Subject Factors in the PCA Recognition of Human Faces," in *CVPR 2003 Workshop on Statistical Analysis in Computer Vision*, 2003.

[312] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper, "The CSU Face Identification Evaluation System: Its Purpose, Features and Structure," *International Conference on Vision Systems*, pp. 304-311, 2003.

[313] R. Beveridge, D. Bolme, M. Teixeira, and B. Draper, "The CSU Face Identification Evaluation System User's Guide: Version 5.0," Computer Science Department, Colorado State University May 1 2003.

[314] M. Kirby, *Dimensionality Reduction and Pattern Analysis: an Empirical Approach.*: Wiley, 2000.

[315] J. Kittler, M. Hatef, D. Robert, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226-239, 1998.

[316] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, pp. 559-572, 1901.

[317] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417-441, 1933.

[318] "Viisage, Inc., Viisage face recognition technology, 2001. Available from <http://www.viisage.com/technology.htm>."

[319] S. C. Chen, Y. L. Zhu, D. Q. Zhang, and J. Y. Yang, "Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA," *Pattern Recognition Letters*, vol. 26, pp. 1157–1167, 2005.

[320] W. S. Yambor, B. A. Draper, and J. R. Beveridge, "Analyzing PCA-based face recognition algorithms: eigenvectors selection and distance measures," in *Empirical Evaluation Techniques in Computer Vision*, H. Christensen and J. Phillips, Eds. Singapore: World Scientific Press, 2000.

[321] A. Eleyan and H. Demirel, "Face recognition system based on PCA and feedforward neural networks," *Lecture Notes in Computer Science*, vol. 3512, pp. 935-942, 2005.

[322] A. Pentland, T. Starner, N. Etcoff, N. Masoiu, O. Oliyide, and M. Turk, "Experiments with Eigenfaces," in *Looking at people, workshop of International Joint Conferences on Artificial Intelligence*. Acapulco, Mexico, 1993.

[323] M.-H. Yang, N. Ahuja, and D. Kriegman, "Face Recognition Using Kernel Eigenfaces," in *Proceedings of 2000 International Conference on Image Processing*, vol. 1. Vancouver, BC, Canada: IEEE Computer Society, 2000, pp. 37-40.

[324] D. Zhou and X. Yang, "Face Recognition Using Direct-Weighted LDA," in *PRICAI 2004: Trends in Artificial Intelligence*, vol. 3157, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 760-768.

[325] S. Shan, P. Yang, X. Chen, W.Gao, and L. AMFG 2005, pp., 2005, "AdaBoost Gabor Fisher Classifier for Face Recognition," in *Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, vol. 3723, *Lecture Notes in Computer Science*, 2005, pp. 278-291.

[326] H. Bischof, H. Wildenauer, and A. Leonardis, "Illumination insensitive recognition using eigenimages," *Computer Vision and Image Understanding*, vol. 95, pp. 86-104, 2004.

[327] R. Epstein, P. Hallinan, and A. L. Yuille, "Eigenimages suffice: An empirical investigation of low-dimensional lighting models," *IEEE workshop on physics-based modeling in computer vision*, pp. 108-116, 1995.

[328] P. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions," in *IEEE conference on computer vision and pattern recognition*. Seattle, WA, 1994, pp. 995–999.

[329] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of a lambertian object," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1322–1333, 2002.

[330] L. Zhao and Y. Yang, "Theoretical analysis of illumination in PCA based vision systems," *Pattern Recognition*, vol. 32, pp. 547-564, 1999.

[331] Y. Xu, D. Zhang, J.-Y. Yang, Z. Jing, and M. Li, "Efficient KPCA-based feature extraction: A novel algorithm and experiments," *Lecture Notes in Control and Information Sciences*, vol. 345, pp. 220-229, 2006.

[332] D. Q. Zhang, S. C. Chen, and Z. H. Zhou, "Recognizing face or object from a single image: Linear vs. kernel methods on 2D patterns," in *Structural, Syntactic, and Statistical Pattern Recognition, Joint IAPR International Workshops, SSPR 2006 and SPR 2006*, vol. 4109, *Lecture Notes in Computer Science*. Hong Kong, China, 2006, pp. 889-897.

[333] M. S. Park, J. H. Na, and J. Y. Choi, "Feature Extraction Using Class-Augmented Principal Component Analysis (CA-PCA)," in *16th International Conference on Artificial Neural Networks, Proceedings, Part II*, vol. 4132, *Lecture Notes in Computer Science*, S. D. Kollias, A. Stafylopatis, W. Duch, and E. Oja, Eds. Athens, Greece, 2006, pp. 606-615.

[334] S. C. Chen and T. K. Sun, "Class-information-incorporated principal component analysis," *Neurocomputing*, vol. 69, pp. 216-223, 2005.

[335] E. P. Vivek and N. Sudha, "Robust Hausdorff distance measure for face recognition," *Pattern Recognition*, vol. 40, pp. 431-442, 2007.

[336] Z. L. Zheng, J. Yang, and M. T. Zhu, "Face detection and recognition using colour sequential images," *Journal of Research and Practice in Information Technology*, vol. 38, pp. 135-149, 2006.

[337] Y. H. He, L. Zhao, and C. R. Zou, "Face recognition based on PCA/KPCA plus CCA," *Lecture Notes in Computer Science*, vol. 3611, pp. 71-74, 2005.

[338] R. Gottumukkal, H. T. Ngo, and V. K. Asari, "Multi-lane architecture for eigenface based real-time face recognition," *Microprocessors and Microsystems*, vol. 30, pp. 216-224, 2006.

[339] H. Y. Zhang, H. D. Ma, and A. L. Ming, "Grid-based multi-scale PCA method for face recognition in the large face database," in *Advanced Web and Network Technologies, and*

Applications, APWeb 2006 International Workshops: XRA, IWSN, MEGA, and ICSE, vol. 3842. Harbin, China: Springer, 2006, pp. 1033-1040.

[340] A. P. Kumar, V. Kamakoti, and S. Das, "System-on-programmable-chip implementation for on-line face recognition," *Pattern Recognition Letters*, vol. 28, pp. 342-349, 2007.

[341] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, pp. 131–137, 2004.

[342] "The ORL Database of Faces. AT&T Laboratories Cambridge. Available from http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html."

[343] A. M. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report #24, June 1998.

[344] "The Yale Face Database. Available from http://cvc.yale.edu/projects/yalefaces/yalefaces.html."

[345] P. C. Yuen and J. H. Lai, "Face representation using independent component analysis," *Pattern Recognition*, vol. 35, pp. 1247-1257, 2002.

[346] M.-H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," in *IEEE Conference on Automatic Face and Gesture Recognition*, 2002, pp. 215-220.

[347] D. Q. Zhang and Z. H. Zhou, "(2D)(2)PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing*, vol. 69, pp. 224-231, 2005.

[348] H. Kong, L. Wang, E. K. Teoh, X. Li, J.-G. Wang, and R. Venkateswarlu, "Generalized 2D principal component analysis for face image representation and recognition," *Neural Networks*, vol. 18, pp. 585-594, 2005.

[349] D. B. Graham and N. M. Allinson, "Characterizing Virtual Eigensignatures for General Purpose Face Recognition," in *Face Recognition: From Theory to Applications*, vol. 163, *NATO ASI Series F, Computer and Systems Sciences*, 1998, pp. 446-456.

[350] P. S. Hiremath and C. J. Prabhakar, "Face recognition technique using symbolic PCA method," in *Proc. Int. Conf. on Pattern Recognition and Machine Intelligence*, vol. 3776. Kolkata, India: Springer Verlag, 2005, pp. 266-271.

[351] D. Choukria and P. Cazes, "Extension of the principal component analysis to interval data," in *NTTS'95: New Techniques and Technologies for statistics*. Bonn, 1995.

[352] W.-Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, pp. 399-458, 2003.

[353] T.-K. Kim, H. Kim, W. Hwang, and J. Kittler, "Component-based LDA face description for image retrieval and MPEG-7 standardisation," *Image and Vision Computing*, vol. 23, pp. 631-642, 2005.

[354] F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovac, "Color-based face detection in the "15 seconds of fame" art installation," in *Mirage 2003, Conference on Computer Vision / Computer Graphics Collaboration for Model-based Imaging, Rendering, image Analysis and Graphical special Effects*, W. Philips, Ed. INRIA Rocquencourt, France, 2003, pp. 38-47.

[355] B. Achermann and H. Bunke, "Combination of Face Classifiers for Person Identification," in *13th International Conference on Pattern Recognition*, vol. 3. Vienna, Austria, 1996, pp. 416.

[356] T. Ho, J. Hull, and S. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 66-75, 1994.

[357] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing*, vol. 11, pp. 467-476, 2002.

[358] A. Timo, H. Abdenour, and P. Matti, "Face Recognition with Local Binary Patterns," in *European Conference on Computer Vision*, vol. 3021, *Lecture Notes in Computer Science*, 2004, pp. 469-481.

[359] W. Zhang, S. Shan, W. Gao, and X. Chen, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition," *IEEE International Conference on Computer Vision And Image Understanding*, vol. 1, pp. 786-791, 2005.

[360] D. Xi, I. T. Podolak, and S.-W. Lee, "Facial Component Extraction and Face Recognition with Support Vector Machines," in *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. Washington, D.C., USA: IEEE Computer Society, 2002, pp. 76-81.

[361] J. Gallant, J. L. Braun, and D. VanEssen, "Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex," *Science*, vol. 259, pp. 100-103, 1993.

[362] P. Fox, J. Cheng, and J. Lu, "Theory and experiment of Fourier-Bessel field calculation and tuning of a pulsed wave annular array," *Journal of the Acoustical Society of America*, vol. 113, pp. 2412–2423, 2003.

[363] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A-Optics Image Science and Vision*, vol. 14, pp. 1724–1733, 1997.

[364] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, pp. 1771–1782, 2000.

[365] A. Lanitis, "Person identification from heavily occluded face images," in *Proceedings of the 2004 ACM Symposium on Applied Computing*, 2004, pp. 5-9.

[366] A. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 748–762, 2002.

[367] A. V. Nefian and M. H. Hayes III, "An embedded HMM-based approach for face detection and recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 6, 1999, pp. 3553–3556.

[368] L. Rabiner and B. Huang, *Fundamentals of Speech Recognition*. NJ: Prentice-Hall, 1993.

[369] S. S. Kuo and O. E. Agazzi, "Keyword Spotting in Poorly Printed Documents using Pseudo 2-D Hidden Markov Models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, pp. 842-848, 1994.

[370] A. V. Nefian and M. H. Hayes, "A Hidden Markov Model for Face Recognition," in *ICASSP 98*, vol. 5, 1998, pp. 2721-2724.

[371] F. S. Samaria and S. Young, "HMM-Based Architecture for Face Identification," *Image and Computer Vision*, vol. 12, pp. 537-543, 1994.

[372] A. L. Yarbus, *Eye movement and vision*. New York: Plenum Press, 1967.

[373] B. Heisele and T. Koshizen, "Components for Face Recognition," in *Conference on Automatic Face and Gesture Recognition*. Seoul, Korea, 2004, pp. 153-158.

[374] B. J. Balas and P. Sinha, "Region-based representations for face recognition," *ACM Trans. Appl. Percept.*, vol. 3, pp. 354-375, 2006.

[375] V. Cantoni and M. Ferretti, *Pyramidal Architectures for Computer Vision*. New York: Plenum Press, 1994.

[376] J. Jolion and A. Rosenfeld, *A Pyramid Framework for Early Vision: Multiresolutional Computer Vision*: Kluwer Academic Publishers, 1994.

[377] A. Rosenfeld and G. J. Vanderbrug, "Coarse-fine template matching," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 2, pp. 104–107, 1977.

[378] B. Heisele, T. Serre, S. Prentice, and T. Poggio, "Hierarchical Classification and Feature Reduction for Fast Face Detection with Support Vector Machines," *Pattern Recognition*, vol. 36, pp. 2007-2017, 2003.

[379] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake, "Efficient face detection by a cascaded support-vector machine expansion," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 460, pp. 3283 - 3297, 2004.

[380] M. Gharavi-Alkhansari, "A fast globally optimal algorithm for template matching using low-resolution pruning," *IEEE Transactions on Image Processing*, vol. 10, pp. 526-533, 2001.

[381] J. You, W. P. Zhu, H. A. Cohen, and E. Pissaloux, "Fast object recognition by parallel image matching on a distributed system," in *7th IEEE Symposium on Parallel and Distributeed Processing*, 1995, pp. 78.

[382] J. MacLean and J. Tsotsos, "Fast pattern recognition using gradient-descent search in an image pyramid," in *Proceedings of 15th International Conference on Pattern Recognition*, vol. 2. Barcelona, 2000, pp. 873 - 877.

[383] J. Shepherd, G. Davies, and H. Ellis, "Studies of cue saliency. In Perceiving and remembering faces," G. Davies, H. Ellis, and J. Shepherd, Eds. London: Academic Press, 1981, pp. 105-131.

[384] I. N. Jarudi and P. Sinha, "Relative Contributions of Internal and External Features to Face Recognition," Massachusetts Institute of Technology, Cambridge, MA CBCL Paper #225/AI Memo #2003-004, March 2003 2003.

[385] N. D. Haig, "Exploring recognition with interchanged facial features," *Perception*, vol. 15, pp. 235-247, 1986.

[386] I. H. Fraser, G. L. Craig, and D. M. Parker, "Reaction time measures of feature saliency in schematic faces," *Perception*, vol. 19, pp. 661-673, 1990.

[387] G. Davies, H. Ellis, and J. Shepherd, "Cue saliency in faces as assessed by the `Photofit' technique'," *Perception*, vol. 6, pp. 263-269, 1977.

[388] P. Nagabhushan, D. S. Guru, and B. H. Shekar, "Visual learning and recognition of 3D objects using two-dimensional principal component analysis: A robust and an efficient approach," *Pattern Recognition*, vol. 39, pp. 721-725, 2006.

[389] L. Wang, T. N. Tan, W. M. Hu, and H. Z. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Trans. Image Proc.*, vol. 12, pp. 1120–1131, 2003.

[390] I. R. Vega and S. Sarkar, "Statistical motion model based on the change of feature relationships: Human gait-based recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1323–1328, 2003.

[391] R. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. IEEE International Conference on Face and Gesture Recognition*, 2002, pp. 366–371.

[392] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon, "A Floor Sensor System for Gait Recognition," in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, 2005, pp. 171-176.

[393] A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.

[394] A. K. Jain, L. Hong, and Y. Kulkarni, "A multimodal biometric system using fingerprints, face and speech," in *2nd International Conference on Audio and Video-based Biometric Person Authentication*, 1999, pp. 182–187.

[395] Y. Wang, T. Tan, and A. K. Jain, "Combining face and iris biometrics for identity verification," in *International Conference on Audio- and Video-Based Biometric Person Authentication*, 2003, pp. 805–813.

[396] L. Hong, A. Jain, and S. Pankanti, "Can multibiometrics improve performance?" in *IEEE Workshop on Identification of Advanced Technologies*, 1999, pp. 59–64.

[397] B. Schiele, "How many classifiers do I need?" in *International Conference on Pattern Recognition*, vol. 2, 2002, pp. 176–179.

[398] A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.

[399] Z. Liu and S. Sarkar, "Outdoor recognition at a distance by fusing gait and face," *Image Vision Comput.*, vol. 25, pp. 817-832, 2007.

[400] A. Kale, A. R. Chowdhury, and R. Chellappa, "Fusion of Gait and Face for Human Identification," *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Processing (ICASSP '04)*, vol. 5, pp. 901-904, 2004.

[401] X. Zhou and B. Bhanu, "Feature Fusion of Face and Gait for Human Recognition at a Distance in Video," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 529-532.

[402] X. Zhou, B. Bhanu, and J. Han, "Human Recognition at a Distance in Video by Integrating Face Profile and Gait," in *5th International Conference on Audio- and Video-Based Biometric Person Authentication*, 2005, pp. 533-543.

[403] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, pp. 324–335, 1993.

[404] B. Bhanu and X. L. Zhou, "Face recognition from face profile using dynamic time warping," in *17th International Conference on Pattern Recogntion*, vol. 4, 2004, pp. 499–502.

[405] X. Zhou and B. Bhanu, "Integrating Face and Gait for Human Recognition at a Distance in Video," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 37, pp. 1119-1137, 2007.

[406] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. Bowyer, "The humanID gait challenge problem: Data sets, performance and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 162–177, 2005.

[407] A. Kale, N. Cuntoor, A. N. Rajagopalan, B. Yegnanarayana, and R. Chellappa, "Gait analysis for human identification," in *Proceedings of 3rd International Conference on Audio and Video Based Person Authentication (AVBPA '03)*, 2003.

[408] S. Zhou and R. Chellappa, "Probabilistic human recognition from video," in *Proceedings of the 7th European Conference on Computer Vision-Part III (ECCV '02)*, vol. 2352, *Lecture Notes In Computer Science*: Springer-Verlag, 2002, pp. 681-697.

[409] G. Shakhnarovich and T. Darrell, "On Probabilistic Combination of Face and Gait Cues for Identification," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. Washington, DC: IEEE Computer Society, 2002, pp. 169-174.

[410] T. K. M. Lee, S. Ranganath, and S. Sanei, "Fusion of Chaotic Measure Into a New Hybrid Face-Gait System for Human Recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*. Washington, DC: IEEE Computer Society, 2006, pp. 541-544.

[411] M. T. Rosenstein, J. J. Collins, and C. J. D. Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D*, vol. 65, pp. 117-134, 1993.

[412] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. USA: McGraw-Hill, 1978.

[413] L. Torres, "Is there any hope for face recognition?" presented at Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004), Lisboa, Portugal, 2004.

[414] R. Jafri and H. R. Arabnia, "PCA-Based Methods for Face Recognition," in *The 2007 International Conference on Security and Management (SAM'07)*. Las Vegas, USA, 2007, pp. 534-541.

[415] L. Lee, "Gait dynamics for recognition and classification," MIT AI Lab, Technical Report AIM-2001-019, September 2001.

[416] S. L. Delp and J. P. Loan, "A graphics-based software system to develop and analyze models of muscoskeletal structures," *Comput. Bid. Med.*, vol. 25, pp. 21-34, 1995.

[417] C. L. Vaughan, G. N. Murphy, and L. L. D. Toit, *Biomechanics of Human Gait: An Annotated Bibliography*: Human Kinetics Pub, 1987.

[418] H. J. Woltring, "A FORTAN package for generalized cross-validatory spline smoothing and differentation," *Advances in Engineering Software*, vol. 8, pp. 104-113, 1986.

[419] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Chapter 3: Interpolation and extrapolation," in *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed: Cambridge University Press, 1992, pp. 994.

[420] D. K. Wagg and M. S. Nixon, "On Automated Model-Based Extraction and Analysis of Gait," in *6th International Conference on Automatic Face and Gesture Recognition*. Seoul, South Korea, 2004.

[421] M. S. Nixon, T. Tan, and R. Chellappa, *Human Identification Based on Gait*, vol. 4. New York: Springer-Verlag, Inc., 2006.

[422] R. Tanawongsuwan and A. Bobick, "Gait Recognition from Time-Normalized Joint-Angle Trajectories in the Walking Plane," in *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, vol. 2, 2001, pp. 726-731.

[423] J.-H. Yoo and M. S. Nixon, "Feature Extraction and Selection for Recognizing Humans by Their Gait," in *International Symposium on Visual Computing*, vol. 2, 2006, pp. 156-165.

[424] D. Winter, "Biomechanics and Motor Control of Human Movement," 2nd ed: Wiley-Interscience, 1990, pp. 277.

[425] A. Y. Johnson, J. Sun, Bobick, and J. A.F., "Predicting large population data cumulative match characteristic performance from small population data," in *Proc. Audio- and Video-based Biometric Person Authentication*. Guildford, UK, 2003, pp. 821-829.

[426] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, pp. 131-156, 1997.

[427] S.-B. Cho and F. E. Pollick, "Toward Person Authentication with Point Light Display Using Neural Network Ensembles," in *Audio- and Video-Based Biometric Person Authentication (AVBPA '03)*: Springer Berlin / Heidelberg, 2003, pp. 878-885.

[428] J. W. Davis and H. Gao, "Gender Recognition from Walking Movements using Adaptive Three-Mode PCA," in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '04)*, vol. 1. Washington, DC: IEEE Computer Society, 2004, pp. 9.

[429] W. L. Wu, F. C. Su, Y. M. Cheng, and Y. L. Chou, "Potential of the genetic algorithm neural network in the assessment of gait patterns in ankle arthrodesis," *Annals of Biomedical Engineering*, vol. 29, pp. 83-91, 2001.

[430] R. Headon and R. Curwen, "Recognizing movements from the ground reaction force," in *Proceedings of the 2001 Workshop on Perceptive User interfaces*, vol. 15. Orlando, Florida: ACM Press, New York, NY, 2001, pp. 1-8.

[431] T. Murakita, T. Ikeda, and H. Ishiguro, "Human tracking using floor sensors based on the Markov Monte Carlo method," in *Proceedings of 17th International Conference on Pattern Recognition (ICPR)*, vol. 4, 2004, pp. 917 – 920.

[432] M. Addlesee, A. Jones, F. Livesey, and F. Samaria, "The ORL Active Floor," *IEEE Personal Communications*, pp. 35-41, 1997.

[433] R. J. Orr and G. D. Abowd, "The smart floor: a mechanism for natural user identification and tracking," in *CHI '00 Extended Abstracts on Human Factors in Computing Systems*. The Hague, The Netherlands: ACM Press, New York, NY, 2000, pp. 275-276.

[434] P. Cattin, D. Zlatnik, and R. Borer, "Sensor Fusion for a Biometric System Using Gait," in *Proc. Multisensor Fusion and Integration for Intelligent Systems (MFI)*. Germany, 2001, pp. 233- 238.