HERIOT-WATT UNIVERSITY

Department of Computer Science

School of Mathematical and Computer Sciences

Recognizing Complex Faces and Gaits Via Novel Probabilistic Models

Ibrahim Venkat@Krishnamurthy Venkatasubramanian

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

October 2010

Abstract

In the field of computer vision, developing automated systems to recognize people under unconstrained scenarios is a partially solved problem. In unconstrained scenarios a number of common variations and complexities such as occlusion, illumination, cluttered background and so on impose vast uncertainty to the recognition process. Among the various biometrics that have been emerging recently, this dissertation focus on two of them namely face and gait recognition.

Firstly we address the problem of recognizing faces with major occlusions amidst other variations such as pose, scale, expression and illumination using a novel *PRObabilistic Component based Interpretation Model (PROCIM)* inspired by key psychophysical principles that are closely related to reasoning under uncertainty. The model basically employs *Bayesian Networks* to establish, learn, interpret and exploit intrinsic similarity mappings from the face domain. Then, by incorporating efficient inference strategies, robust decisions are made for successfully recognizing faces under uncertainty. PROCIM reports improved recognition rates over recent approaches.

Secondly we address the newly upcoming gait recognition problem and show that PROCIM can be easily adapted to the gait domain as well. We scientifically define and formulate *sub-gaits* and propose a novel modular training scheme to efficiently learn subtle sub-gait characteristics from the gait domain. Our results show that the proposed model is robust to several uncertainties and yields significant recognition performance. Apart from PROCIM, finally we show how a simple component based gait reasoning can be coherently modeled using the recently prominent *Markov Logic Networks (MLNs)* by intuitively fusing imaging, logic and graphs.

We have discovered that face and gait domains exhibit interesting similarity mappings between object entities and their components. We have proposed intuitive probabilistic methods to model these mappings to perform recognition under various uncertainty elements. Extensive experimental validations justifies the robustness of the proposed methods over the state-of-the-art techniques.

Acknowledgements

I would like to express my sincere thanks to my Ph.D. guide Prof.Philippe De Wilde who played the roles of a friend, philosopher and guide and continuously supported me both during hard and happiest moments of my research progress. I thank Heriot-Watt University for offering me the James-Watt Scholarship, to pursue my studies. I thank TATiUC, Malaysia for providing me partial financial support during my Ph.D. studies. I thank Prof.Shukri, Dr.Tajudin and Prof.Rosni of USM, Malaysia for offering me career guidance. I wish to thank Prof.Mike Chantler for providing me valuable advice during the internal vivas. I thank Prof.David Corne, the director or research, who provided me plenty of opportunities to participate in numerous exhibitions, promotions and present various seminars.

I thank June Maxwell for arranging necessary Ph.D. meetings and providing scribe opportunities regularly. I whole heartedly thank the help desk team, Donald, Iain, Steve, Adrian and Susan, for providing prompt technical services. My sincere thanks to Claire Porter who provided time to time advice on post graduate procedures. I thank Christine McBride for providing all the travel assistance to conferences and other school events. I thank the school office staff, Elizabeth and her friends for providing all the administrative assistance.

I am very grateful to my beloved wife Shahida, who took unpaid leave from her service and supported me throughout my studies with a smiling face. Also I thank my aged mother, who traveled all the way from India to Edinburgh and stayed with me to give all the moral support. I thank my brother, brother-in-laws, sister-in-laws and all my relatives for their encouragement and support during my studies. My thanks are due to other fellow students, Festus, Xiobin, Ali, Sarah, Pratik, Amol and many more, who shared their valuable experience, joys and sorrows.

Thanks to everyone who took the time to read through parts of my thesis and provided valuable suggestions (Prof. Rajaraman, Dr.Farid, Dr.Rehan, Dr.Yusof and Dr.Osama). I thank scientists Dr.Rajasankar and Dr.Nagesh of CSIR and friends of TATA group for providing all the necessary technical advice.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Face Recognition

Biometrics is the science of automated methods for identifying people and a biometric is a distinctive physiological (e.g., face, fingerprint, iris) or behavioral (e.g., gait, signature, voice) characteristic that can be used to identify a person [14]. Biometrics-enabled intelligence has rapidly become an accepted tool for solving immediate identity problems as biometric identity data is readily indexed, sorted and stored [15]. Intelligence analysts and law enforcement personnel use biometric identity data as the central criteria to establish identity and as a basis to recommend action. A key objective of machine vision researchers is to build automated recognition systems that can compete and eventually surpass human visual intelligence. A recent comprehensive survey [1] shown in Fig.1.1 depicts that annual revenues from biometric industries are estimated to multiply about three times by the year 2014, when compared to that of 2009. This motivates the fact that the field of biometrics has a promising future.

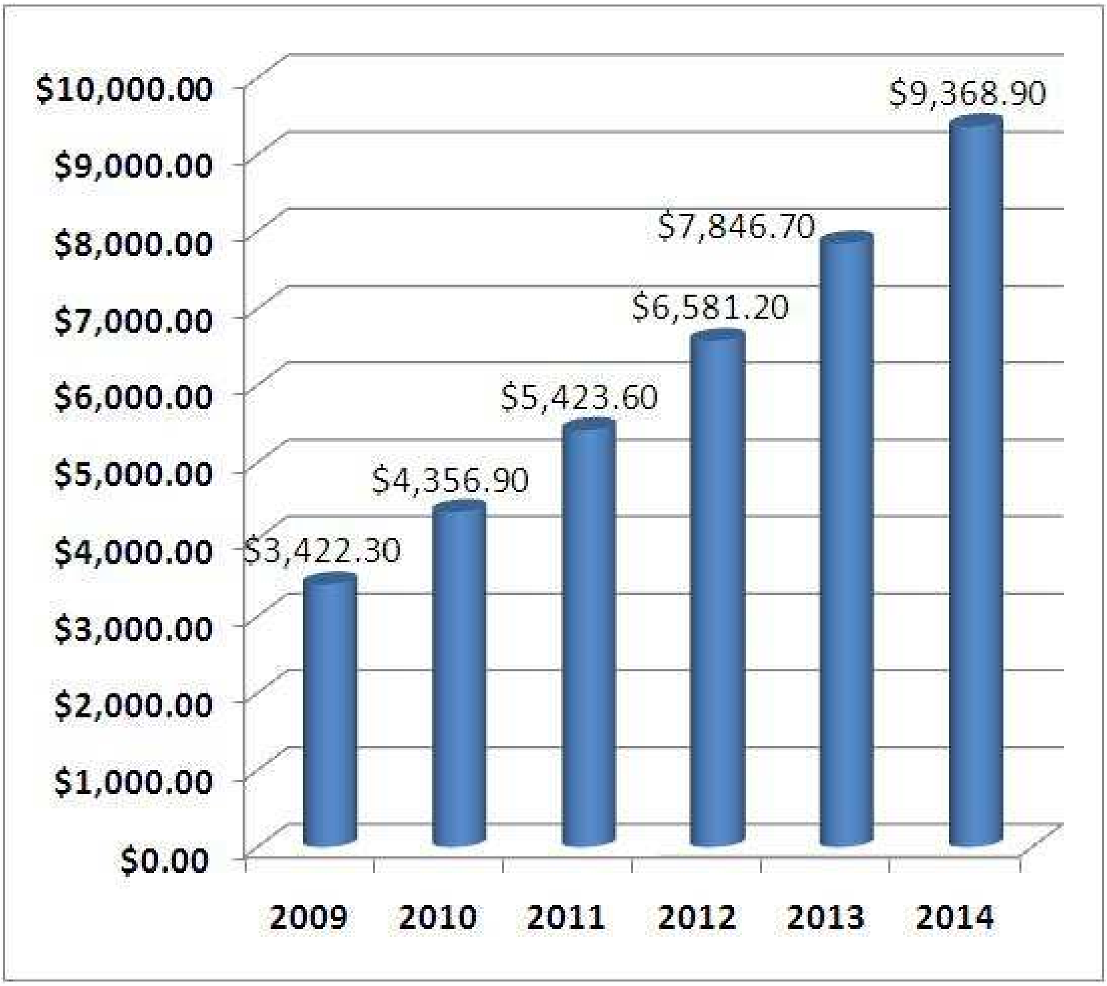


FIGURE 1.1: Annual biometric industry revenues, 2009-2014 ($m USD), as projected by International Biometrics Group [1].

Briefly the problem of machine recognition of faces can be defined as identifying a set of face images from a stored face database that closely match a given probe face image. In other words, for a given probe face, closely matching faces are rank-listed by a *Face Recognition (FR)* system. *FR* differs from *face detection* and *face verification* though all these terms are related to some extent. Face detection is a technique that determines the locations and sizes of human faces or a face in arbitrary digital images and thereby helps to specifically segment face images or a face from the rest of the background. Face verification systems, on the other hand, authenticate a person's identity by comparing the captured biometric characteristic with the person's own biometric template(s) pre-stored in the system, a one-to-one match to determine whether the identity claimed by the individual is true. Facial biometric information is fused with contextual information to produce useful and actionable intelligence. The technology being nonintrusive in nature has the unique advantage of answering important queries such as:

- How can faces in the crowd be linked to other intelligence information?
- Is the presence of multiple people in the same location an event of interest?
- Can the anticipated presence of an individual be exploited?

Such crucial investigations are often carried out in unconstrained scenarios. While humans have the natural ability to recognize faces in a complex scenario, the problem of recognizing faces amidst uncertainty elements such as occlusion and noise is still only a partially solved problem in machine vision research. Two-dimensional (2-D) FR is of interest in many verification and identification applications such as crowd surveillance and access control [16, 17]. It has been widely used in critical real world applications such as biometric passports, voting polls, criminal investigation and so on. It is gradually becoming part and parcel of consumer products and getting embedded into cars, ATMs, and mobile devices. Digital image organizers such as *Google's Picasa* and *Apple's iPhoto* are typical examples of popular commercial applications which deploy face recognition technology. The pie chart shown in Fig.1.2 clearly shows that FR systems have a market share next to fingerprint technology [1]. Consumer industries which produce biometric featured products would like to make their customers at ease without forcing constraints on them, such as the removal of facial accessories (eg. sunglasses or cap). Moreover it is sensitive or against the norms and practice of certain communities to remove head covers for the need of adhering to the limitations of a technology. Hence research on recognizing faces in unconstrained scenarios, specifically occlusion invariant FR approaches plays a vital role to not only aid the reduction of crime rates, but also to protect social and human rights.

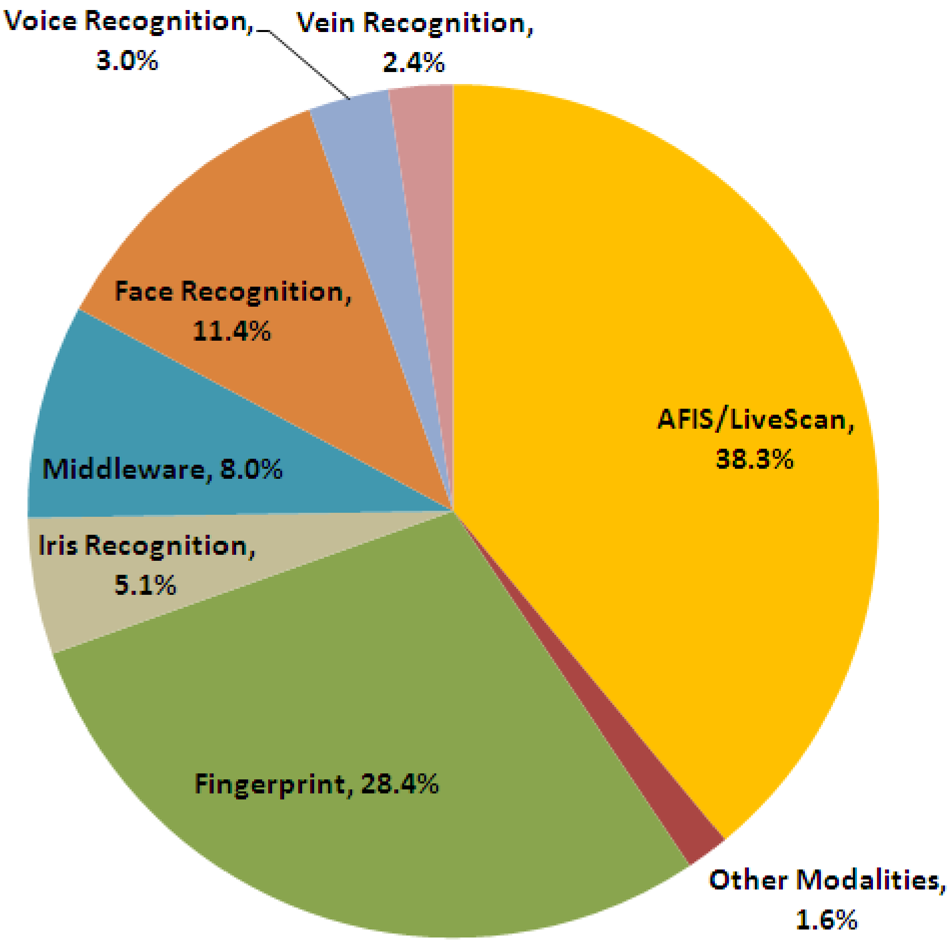


Figure 1.2: The trend of biometric market for various biometric technologies as reported by International Biometrics Group [1].

1.2 Gait Recognition

The literal meaning of gait is “manner of walking” [18] and the oldest gait analysis is due to Aristotle [19]. The notion of using gait to recognize people has been even mentioned by Shakespeare “For that John Mortimer...in face, in gait in speech he doth resemble” (Henry IV/II) [20]. An evidence from biomechanics literature [21] states that “A given person will perform his or her walking pattern in a fairly repeatable and characteristic way, sufficiently unique that it is possible to recognize that person at a distance using his/her gait”. The problem of identifying individuals at a distance by observing their walking behavior when other features such as face and hand geometry are not clearly visible is a common task for the human visual system. Applying computer vision techniques to automate this task is formally known as Gait Recognition (GR).

Progress on this newly emerging biometric has been mainly initiated and contributed by *Defense Advanced Research Projects Agengy's (DARPA's) Human ID at a Distance* program led by Dr.Jonathon Phillips from *National Institute of Standards in Technology (NIST).* The DARPA program focused on three main biometrics viz., face, gait and new technologies. The HumanID gait challenge problem provides a scientific basis for advancing and understanding automatic gait recognition and processing. It offers a challenging data set, a set of experiments in the increasing order of difficulty and a baseline algorithm. Purposefully, DARPA's gait challenge dataset was captured outdoors to force the development of computer vision based gait algorithms to handle several uncertainty issues. Algorithms have to handle complications generated from a person's shadow from sunlight and moving background. Factors that can affect a person's gait in outdoor settings include surface type, shoe-wear type, weight carried and clothing. Video data of gait is also dependent on the viewpoint. It is important to understand the ability of gait recognition algorithms in the presence of these variations. Further, gait recognition algorithms often perform poorly because of low resolution video sequences and subjective human motion. Despite these challenges, gait recognition research is gaining momentum due to increasing demand and more possibilities for deployment by the surveillance industry. Therefore every research contribution which significantly improves this new biometric is a milestone.

1.3 Why a Probabilistic Model?

The underlying problem of recognizing faces and gaits in unconstrained scenarios, obviously involves reasoning under uncertainty. Probability theory which acts as a pivotal guidance to probabilistic models could be regarded as an extension of Boolean logic to situations involving uncertainty [22]. The use of probabilistic models to represent uncertainty, however, is not a matter of ad-hoc choice, but is inevitable if we are to respect common sense while making rational coherent inferences. Probability theory provides a consistent framework for the quantification and manipulation of uncertainty and hence forms one of the central foundations for object recognition [23]. This consistency makes a probabilistic model feasible to successfully cope up with uncertainties in the data, image preprocessing steps such as alignment and normalization, data compression as a consequence of feature extraction and certain degrees of approximation associated with the ultimate decision making process. The intricacies of the uncertain scenario where we intend to capture faces and gaits imposes the reasoning system to consider different possibilities. A conventional approach usually considers any state of the world that is possible and could simply list all the possible outcomes without assigning any priorities. In order to make meaningful decisions, a recognition system need to reason not just about what is possible, but also about what is probable. The calculus of probability theory provides us with a formal framework for considering multiple possible outcomes and their likelihood. It defines a set of mutually exclusive and exhaustive possibilities and associates each of them with a probability. The liberating nature of probabilistic models allow us to make this fact explicit and therefore provide a model which is more faithful to reality [24].

Primarily we will attack the problems using a prominent probabilistic graphical model called Bayesian Network(BN). A BN (or a belief network) is a probabilistic graphical model that represents a set of variables and their conditional independencies via a Directed Acyclic Graph (DAG). Bayesian models have been applied in various applications that work in unconstrained and realistic environments and they are now the mainstay of the AI research field known as "reasoning under uncertainty" [25]. Finally we will also provide a basic framework using a recently growing technique called *Markov Logic Network* by fusing the imaging domain, first-order logic and probabilistic graphical models.

1.4 The Problem

Broadly this dissertation will focus on two vital problems in the field of biometrics namely face and gait recognition under challenging real world scenarios. However we will precisely outline the problems we undertake as follows:

Notwithstanding the enormous research effort that has been invested about two decades to solve the problem of automatic face recognition, we are yet to witness a robust face recognition system that can recognize faces with major occlusions coupled with other uncertainty challenges thrown by the various intricacies in an unconstrained environment. The only system that can successfully recognize faces in a complex scenario is the human visual system which is complimented by two eyes, the so called biological cameras and the extraordinary brain.

i) Hence it makes eminent sense as a first step to extend the research beyond the pixel domain to investigate and identify "What are the key cognitive psychology principles that governs the human visual system and enhances visual intelligence in order to tackle this potential problem?".

ii) Further we will investigate "While popular face recognition techniques such as [26, 27] can successfully recognize faces in controlled conditions, Why are they not feasible to recognize faces in unconstrained scenarios?".

iii) Importantly how to scientifically transform the phenomenological human intuitions and associated key psychological principles (identified in (i)) into a robust probabilistic FR model and validate it with defacto standard experiments and datasets.

We have given the literal meaning and described the problem of gait recognition in general terms in the introductory chapter 1.2. Scientifically the objective of a gait recognition algorithm is to apply image processing and statistical pattern recognition techniques to find who among the humans in a database of gait video sequences closely resemble the gait of a given gait video sequence. Though the problem has been stated in simple terms, it is quite challenging for a machine vision researcher to address this bulk video processing task using normal *Personal Computers*.

iv) Many of the existing gait recognition techniques, manually label several parts of a gait sequence such as head, torso and arms and then propose various algorithms to represent these parts and process them to classify gaits. How can we formulate an approach which will avoid such manual labeling but at the same time extract and exploit useful information from various gait components?

v) How can we design and incorporate a machine learning procedure to learn subtle gait characteristics from the abstract pixel domain?

vi) Statistical relational techniques aim to attack complexity and uncertainty by unifying two major paradigms namely, logic and probability. Additionally can we formulate a fusion based architecture that can bridge the imaging domain with first order logic and probabilistic graphical models to reason gait recognition?

This dissertation will provide scientific means to answer these interesting research questions.

1.5 Contributions

This dissertation presents novel methods based on probabilistic methods, inspired by key psychophysical principles, to address the problems of recognizing faces and gaits under unconstrained scenarios.

We have systematically synthesized, analyzed and presented an up-to-date overview of various face and gait recognition models and clearly demonstrated where they stand after being progressively evolved and improved over the past few years. We have investigated the strengths and limitations of various approaches. Importantly we have investigated the factors that affect some well known approaches, specifically in unconstrained environments (problem (ii), section 1.4).

We have identified an optimal way of segmenting a face without loosing any information or without any redundancies as a vital income of a thorough literature survey. We also propose an automatic segmentation scheme for the gait domain which does not require manual labeling of components. We apply set theory notations which provide simple but yet efficient mathematical means to formulate the proposed segmentation scheme.

We have defined an important phenomenon which we term "*Influence Strength*" which intuitively quantifies how much strength is inherent in a particular component of an object entity in order to influence the recognition mechanism of the object itself. We have shown that this general phenomenon leads us to hypothesize that *"The face or gait pattern of the probe object being recognized by observing the pattern of its component will be more similar to the corresponding gallery pattern, if the magnitude of influence strength is high"*.

This dissertation also contributes to the discovery of a concept called "*Similarity Mappings*". The proposed **PRO**babilistic **C**omponent **I**nterpretation **M**odel which we abbreviate as (**PROCIM**) is based on a fundamental insight about human pattern matching and memory. While reasoning with objects which are prone to uncertainties, humans are often able to notice similarities between object components and the objects that represent those objects. For example in complex scenarios often the complete face of a subject is not visible due to the presence of occlusions and other uncertainty factors such as pose. However, still humans are often able to notice similarity between particular features or a combination of features from the available cues in the face which can help them to identify (guess) few faces. Similarly when we see a person walking at a distance, we may notice a particular pattern of arm-swinging or hip movement as a characteristic of the whole walking gait of that person. This similarity based reasoning is processed in such a way that it reveals inherent conditional independencies between object entities. We have scientifically represented these independencies using Bayesian Networks(BN) in this dissertation. Basically, here we contribute to address the problems (i) and (iii), stated in section 1.4.

Further this thesis proposes a novel robust formula which will enable *PROCIM* to make meaningful decision under uncertainty by bridging probability theory and utility theory. This formula makes use of the probability distribution as well as influence strengths entailed by the Bayesian Network to successfully classify probes which are prone to complexities.

We have scientifically defined and formulated sub-gaits and various sub-gait operators and modeled a novel modular training scheme which enables *PROCIM* to learn and exploit subtle sub-gait characteristics from the gait domain. The formulation of sub-gaits enable PROCIM to avoid manual labeling of several parts of a gait sequence. This contribution will attack the problems (iv) and (v), which have been stated in section 1.4. Further we have logically shown how to interpret the combination of several sub-gait operators in order to reveal intrinsic characteristics of gait patterns. It has been demonstrated that the proposed modularity based reasoning aids *PROCIM* to mitigate the uncertainties encountered by the objects.

Finally this thesis presents a basic framework to learn gait component relationships by fusing three diverse domains viz., imaging, logic and graphs. This fusion based framework shows, how a simple component based gait reasoning approach can be coherently modeled using a newly upcoming statistical relational technique called Markov Logic Networks. This fusion based contribution will address problem (vi) stated in section 1.4.

1.6 Awards / Publications

* Ibrahim Venkat and Philippe De Wilde. Research proposal - Bayesian Object Recognition Ahead of HuMANs (BORAHMAN). Won a cash award for being finalists of the Thales Scottish Technology Prize, Glasgow, December 2009.

* Ibrahim Venkat and Philippe De Wilde. Learning gait component relationships by fusing logic and graphs using Markov Logic Networks. In S.Maskell and S. Godsill, editors, Proceedings of FUSION2010. IET, London, July 2010.

* Ibrahim Venkat and Philippe De Wilde. Robust Gait Recognition by Learning and Exploiting Sub-gait Characteristics. Intl. Journal of Computer Vision, August 2010 (www.springerlink.com).

* Ibrahim Venkat and Philippe De Wilde. Psychophysically inspired similarity mappings to recognize faces with major occlusions in unconstrained scenarios. Image and Vision Computing, Elsevier, under review.

Previous awards relevant to the Ph.D. problems

* Gold Medal from Ministry of Science,Technology & Innovation, Malaysia,Feb 2006

* Gold Medal from Korean Invention Association, Seoul Intl. Invention Fair, Korea, Dec 2006

Both for the invention and innovation of *Intelliface*, an intelligent face recognition system.

1.7 Dissertation Outline

This dissertation is organized as follows:

Chapter 2 provides the motivation gained from related work in the field of face recognition. Beyond the pixel domain, various research works that address the problem of face recognition from psychology and neuroscience are presented initially. This is followed by an overview of various probabilistic models relevant to the problem. Then literature specific to occluded face recognition is presented. Finally the chapter clearly projects the present state of occlusion models.

Chapter 3 presents the theoretical constructs of the proposed *PROCIM* architecture. Based on the psychophysical findings an efficient segmentation scheme that decomposes the faces into various subregions is presented initially. Then how the phenomenological similarity mapping concept is realized from the facial domain is presented. Further the standard way of learning parameters from the Bayesian Network and a novel formula to make decision under uncertainty is presented. Finally performance evaluation of *PROCIM* using standard datasets and experiments is demonstrated.

Chapter 4 presents an overview of the newly emerging gait recognition algorithm. Initially a brief introduction to gait analysis techniques from diverse fields such as medical, biomechanical and psychological literature is presented. This chapter proposes a general classification scheme to organize various gait recognition algorithms according to the basis of data acquisition and systematically briefs them. Specific emphasis is given to video sensor based gait recognition algorithms.

Chapter 5 basically presents how the *PROCIM* architecture can be extended to address the gait recognition problem. Using set theory notations, definition and formulation of sub-gaits and the design of a modular training scheme are presented in detail. Then the robustness of the proposed *PROCIM* technique against common variations is analyzed. Further this chapter empirically shows how potential sub-gaits are identified from the various possible sub-gaits. Finally *PROCIM* has been experimentally validated and compared against state-of-the-art gait recognizers at the end of this chapter.

Chapter 6 presents a basic fusion based framework to learn gait component relationships. Initially an overview of *Statistical Relational Learning (SRL)* models and related work on the newly upcoming SRL technique called Markov Logic

Networks (MLNs) are presented. Then some basic concepts about first-order logic relevant to the MLN techniques are introduced. This is followed by a briefing about MLNs. Then a three stage architecture by fusing the imaging, logic and graphical layers using MLN is proposed. Finally a comparison of the proposed method with other standard methods is presented at the end.

Chapter 7 finally concludes this thesis by summarizing the contributions and discusses future avenues of our research.

Chapter 2

Motivation from Related Work: Face Recognition

2.1 Inspiration from Cognitive Psychology and NeuroScience

Enhancing machine vision systems with psychophysical mechanisms enables vision systems to take intelligent decisions on a par with the human mind, especially when complexities such as occlusions are encountered. In other words, the twin enterprises of visual neuroscience and computer vision have deeply synergistic objectives and an understanding of human visual processes involved in face recognition can facilitate better computational models [28]. Psychologists are of the view that psychologically feasible computational models exhibit clear and strong relationships between behavior and properties of the domains which they intend to represent [29]. Many successful face recognition and verification approaches [26, 30–33] have derived their fundamental ideas from principles of cognitive psychology and neurophysiology. The earliest work on face recognition can be traced back at least to the 1950s in the field of psychology [34]. Many studies in psychology and neuroscience have direct relevance to engineers interested in designing algorithms or systems for machine recognition of faces [16]. For example, findings in psychology [35, 36] about the relative importance of different facial features have been noted in the engineering literature [37]. On the other hand, machine systems provide tools for conducting studies in psychology and neuroscience [38, 39].

Behrmann and Mozer [40] have performed a series of psychological experiments to study how humans process occluded objects. Their study shows that humans, in order to minimize the processing load, organize a complex occluded object into subregions and then attend selectively to particular physical regions. This selective attentional spotlight focuses on areas of interest and facilitates preferential processing of information from those chosen areas. This object-based mechanism, in which complex visual input is parsed into discrete units for further processing, has received considerable empirical, neuropsychological, and computational support in recent years.

Vision is a subfield of cognitive science which involves psychological inferences in the higher nervous system, based on learned models gained from experience [41, 42]. It has been conjectured that the brain learns a generative model of how scene components are put together to explain the visual input and that vision is a process of inference in these models [41]. Among the many observations made by Sinha et al. [28], *"increasing familiarity"* is an important factor which enables humans to recognize highly degraded face images, which they ascertain from [43–46]. Additionally, body structure and gait information are much less useful for identification than facial information, even though the effective resolution in that region is very limited. Recognition performance changes only slightly after obscuring the gait or body, but is affected dramatically when the face is hidden. Even police officers with extensive forensic experience perform poorly unless they are familiar with the target individuals. Psychologists [45, 46] raised a precise fundamental question: How does the facial representation and matching strategy used by the visual system change with increasing familiarity, so as to yield greater tolerance to degradations? Exactly what aspect of the increased experience with a given individual leads to an increase in the robustness of the encoding; is it the greater number of views seen or is it the robustness an epiphenomenon related to some biological limitations such as slow memory consolidation rates? The appropriate benchmark for evaluating machine-based face recognition systems is human performance with familiar faces. From the point of computer vision, we perceive that computationally this familiarity aspect can be achieved by machines using efficient machine learning strategies to familiarize known faces and gaits. This would enable the proposed model to tolerate the uncertainties including processing faces and gaits from low resolution images and videos.

Many psychological investigations [47–53] have been attempting to find the relationships between statistical properties of face images and the underlying psychophysical aspects of human facial processing strategies. Recently it has been empirically shown that facial identity information is conveyed largely via mechanisms tuned to horizontal visual structure of face images [54, 55]. Specifically, humans perform substantially better at identifying faces that have been filtered to contain just horizontal information compared to any other orientation band. Dakin and Watt [54] have further shown that processing faces in terms of horizontal structures has computational advantages. As for visual stimuli, face images reveal a noticeable statistical regularity that comes as an approximately linear decrease of their (logarithmically scaled) amplitude spectra as a function of spatial frequency [56–58]. Numerous psychophysical experiments indicate that the maxima in the amplitude spectra are caused by the compound effect of horizontally oriented internal face features [55, 59]. Goffaux et al. [60] did a comparative analysis between the vertical and horizontal relations of facial features. Their experiments provide clear evidence that inversion dramatically disrupts the ability to extract vertical relations between facial features but not horizontal relations. These psychological findings motivate us to consider processing the face in term of horizontally orientated structures.

Hulme and Zeki [61] have investigated the response of brain neural activity in response to occluded faces. The authors have horizontally traversed the face images with an opaque rectangular block to occlude the faces in their experiments.

Lu and Liu [62] have studied the human recognition memory with respect to objects including occluded faces. They have used randomly distributed rectangles to occlude the face images. Such regular shape occlusions have enabled them to quantify the presence of occlusions. "How faces are perceptually encoded within the human visual system? " This is hypothesized in [63] as follows. Faces are represented by the local shapes of their distinctive features and the spatial relationships among these features. Wallis et al.[64] conclude that facial discrimination tasks performed by the human visual cortex rely on the combination of multiple local feature analyzers rather than global information.

Similarity is a basic concept in cognitive psychology which is utilized to explore the principles of human perception [65]. Recent studies [66, 67] refer to the classical contrast model of similarity [68] which insists that perceived similarity is the result of a feature-matching process. One of the fundamental hypothesis [69] associated with the perception and memory of faces states that, humans perceive and remember faces chiefly by means of facial features. Psychological experiments [69] that evaluate similarity judgments supports this hypothesis. Facial processing algorithms used by popular imaging applications such as *photofit* and *identikit* are based on this cognitive phenomenon.

2.2 Probabilistic Models for Reasoning under Uncertainty

Remarkable progress in mathematics and computer science has led to a revolution in the scope of probabilistic models. An exciting development over the last decade has been the gradually widespread adoption of probabilistic models in many areas of computer vision and pattern recognition. Computational approaches to sorting out plausible explanations of data using Bayes rule were pioneered by Thomas Bayes and Pierre-Simon Laplace in the 18th century, but it was not until the 20th century that these approaches could be applied to vision problems using computers [41]. The availability of computer power motivated researchers to tackle larger problems and develop more efficient algorithms and consequently in the past two decades, we have seen a flurry of intense, exciting, and productive research in complex, large-scale probability models and algorithms for probabilistic inference and learning. Probability theory always had a dual aspect, serving both as a normative theory for correct reasoning about chance events, but also as a descriptive theory of how people reason about uncertainty as providing an analysis, for example, of the mental processes of an intelligent juror [70]. Due to uncertainty elements such as multiple occlusions, same objects can result in the formation of different images, and different objects can result in the formation of similar images. Probabilistic models offer the promise to model natural images such as faces which are often prone to such dual uncertainty [71, 72]. For the problem of image segmentation and image parsing, probabilistic models based on the principle of *"Analysis by synthesis"*, where low-level cues combined with spatial grouping rules activate hypotheses about objects have offered reliable solutions [73, 74]. Interestingly this principle relates to the forward and backward projections in the brain [75–81]. Yuille et al. [82] has treated vision as an inverse inference problem where the goal is to estimate the factors that have generated the image and which of those factors should be estimated. Notably they have applied Bayesian inference to design theories of vision that deal with the complexity of images including faces using recent examples from computer vision.

Krynski and Tenenbaum [83] have proposed Bayesian networks as tools to systematically analyze how humans make judgements under uncertainty. Intille and Bobick [84] have demonstrated how highly structured, multi-person action, prone to multiple sources of visual perceptual uncertainties, can be recognized using a Bayesian framework. Dahyot and Heitz [85] have suggested a Bayesian approach inspired by probabilistic principal component analysis to detect objects subject to cluttered backgrounds coupled with occlusions.

Tong et al. [86] have proposed a Bayesian model to recognize facial expressions when faces are subject to uncertainties such as occlusions, pose and illumination variation. Their model is capable of representing relationships among facial action units with conditional dependence links. The performance of facial action unit recognition algorithms gets affected by errors encountered during the feature extraction and face alignment process due to uncertainties such as occlusions. The authors claim that their model, by exploiting intrinsic relationships among the facial action units, could compensate these errors considerably. Bayesian graphical models have been investigated in a number of face recognition studies[87–90].

Similarity measures play a crucial role in theories of recognition, identification, and categorization of objects, where a common assumption is that the greater the similarity between a pair of objects, the more likely the objects are closer within their feature space. Typical similarity metrics such as the Euclidean distance metric correspond to a standard template-matching approach to address recognition. Contrast to such conventional metrics, Moghaddam et al. [27] introduced a probabilistic similarity measure based on a Bayesian analysis of image differences. This measure is based on the following assumption. The probability that the image-based differences denoted by $d(I1, I2)$, of two face images I_1, I_2, are characteristic of typical variations in appearance of the same object. Moghaddam et al. discovered and exploited two mutually exclusive classes of variations that naturally exist in the facial domain called intra-personal and extra-personal variations. The first one, Ω_I, correspond to variations in the appearance of the same individual, due to factors such as different expressions or pose. The later one denoted by, Ω_E, account for variations that exist between different individuals. The similarity measure $S(I_1, I_2)$ has been defined as

$$S(I_1, I_2) = P(\Omega_I | d(I_1, I_2)), \tag{2.1}$$

where $P(\Omega_I | d(I_1, I_2))$ is the a posteriori probability given by Bayes rule, using estimates of the likelihoods $P(d(I_1, I_2)|\Omega_I)$ and $P(d(I_1, I_2)|\Omega_E)$. These likelihoods have been derived from the training face images using a subspace method for density estimation of high-dimensional data [91]. This probabilistic framework is particularly advantageous in that the intra/extra density estimates explicitly characterize the type of appearance variations which are critical in formulating a meaningful measure of similarity. For example, the differences corresponding to facial expression changes (which may have high error norms) are, in fact, irrelevant when the measure of similarity is to be based on identity. The subspace density estimation method used for representing these classes thus corresponds to a learning method for discovering the principal modes of variation important to the classification task. Furthermore, by equating similarity with the a posteriori probability, an optimal non-linear decision rule for matching and recognition has been obtained which makes the approach significantly unique from methods which use linear discriminant analysis techniques (Eg.[37, 92]) for object recognition.

Vast uncertainty is encountered as a consequence of pose variations and probabilistic approaches have been proven to be good in recognizing and detecting faces amidst pose variations [93–99]. The Bayesian probabilistic approach proposed by Sarfraz and Hellwich [93] initially finds a generative function for several pose variations and then use a view-point discriminative method to model the appearance variations corresponding to each pose explicitly. The goal is to create a model that can predict how a given face will appear when viewed at different poses which seems to be an intuitive formulation for the recognition task especially in unconstrained scenarios. Similarities between extracted features of faces at frontal and all other views have been computed and the distribution of these similarities is then used to obtain the likelihood functions of the form $P(I_g, I_p|C)$, where C refers to classes when the gallery I_g and probe I_p images are similar, S, and dissimilar, D, in terms of a subject's identity. The authors approximate the joint likelihood of a probe and gallery face as

$$P(I_g, I_p|C, \phi_g, \phi_p) \approx P(\gamma_{pg} \mid C, \phi_g, \phi_p) \tag{2.2}$$

where ϕ is the pose angle for the corresponding gallery and probe face and γ_{pg} is the similarity between gallery and probe image. The goal is achieved by learning the approximated joint probability distribution of a gallery and probe image at different poses.

Probabilistic frameworks have also been applied to recognize people from facial video information. Stochastic tracking and recognition approaches are based on a unified probabilistic framework, in which individuals are simultaneously tracked and recognised by estimating the posterior probability density function of a Time Series State Space Model (TSSSM) [100–102]. Tracking is formulated as a Bayesian inference problem, and it is solved as a probability density propagation problem (due to the temporal nature of tracking itself); recognition is obtained by applying the MAP rule on the posterior probabilities.

The idea of developing a generic approach using particle filtering [103] was first introduced by Li and Chellapa [104] for stochastic tracking and verification of humans. They implemented a simplified TSSSM with no identity variable, in which only the tracking motion vector was estimated and propagated. They also proposed two facial representations for the observations: the common intensity images of the face, and an Elastic Graph Matching (EGM) representation of the facial landmarks. Their work unfortunately failed to provide any experimental evaluations. Then, Zhou et al. [105] improved the approach of Li and Chellappa, by including both the tracking motion vector and the identity variable in the TSSSM. They also considered several observation likelihoods, and introduced a more complex one by explicitly modelling: the appearance changes within videos using a truncated Laplacian and the intra-personal appearance variations using a probabilistic subspace density, proposed by Moghaddam [106]. More interestingly, the authors developed a probabilistic learning approach to automatically build user models from video frames.

By deriving an adaptive version, Zhou et al. [107] successfully refined their previous recognition system. They modified the observation likelihood by modeling the appearance changes within videos using an adaptive appearance model, the intra and extra-personal appearance variations using a probabilistic subspace density model[106], and suitably weighting frontal view frames using a different probabilistic subspace density model. Then the authors proposed an adaptive motion model, which consisted of: an adaptive velocity model which is derived using a first-order linear predictor based on the appearance difference between the incoming observation and the previous particle configuration; an adaptive noise component function to compute the prediction error and an adaptive technique to adjust the number of particles based on the degree of uncertainty in the noise component. Further, they included an occlusion handling technique based on robust statistics [108–110] to reduce the influence of outliers on the estimation process.

Apart from TSSSM which aims to simultaneously track and recognize individuals, a novel probabilistic appearance manifold approach has been proposed by Lee et al. [111] which is an extension of the approach introduced by Murase and Nayar [112]. The authors applied Bayesian inference to include the temporal coherence of human motion in the distance calculation; in fact, they replaced the conditional probability by using the joint conditional probabilities, which were recursively estimated using the transitions between sub-manifolds. In the experimental results obtained using a small database (20 individuals), the proposed approach outperformed standard image-based recognition techniques. It showed better robustness and stability than a majority voting strategy or a similar system without temporal coherence. Further the approach was able to detect identity changes and handle large pose variations.

2.3 The Occlusion Challenge for Face Recognition Systems

Recently the importance of occlusion invariant face recognition has received considerable attention by the machine vision community as well as from other fields such as cognitive psychology and neuroscience. Face images are often prone to occlusion coupled with other common variations such as illumination, scale and pose in unconstrained environments. In this section we will investigate briefly the scientific basis, strengths and limitations of core occlusion models reported in the literature. On the basis of processing, FR algorithms can be broadly classified into either holistic or component-based models.

2.3.1 Holistic Models

Holistic approaches which are also known as appearance-based methods process the entire face as a whole entity. In holistic template-matching systems each template could be a prototype face, a face like gray-scale image, or an abstract reduced-dimensional feature vector that has been obtained through processing the face image as a whole [37]. Generally, this category of algorithms operate directly on instances of face objects and processes the images as 2D holistic patterns, avoiding therefore the difficulties associated with 3D modeling and landmark detection. While the traditional image-based approaches require many training face images in order to recognize faces in a variety of viewing conditions, the key aspect of the appearance-based scheme is the use of only a small amount of data (the most representative samples) for recognition, thus leading to low memory requirement and high speed processing [113].

The importance of the occlusion problem has been foreseen and specifically illustrated at the earlier stage of automated face recognition research [26]. Turk and Pentland who demonstrated the first successful automatic face recognition system using the eigenface technique stated that occlusions can gracefully degrade recognition performance. This widely used PCA based technique treats the whole face image as a point in a low dimensional space. Each individual face has been represented as a linear combination of uncorrelated orthogonal components known as eigenfaces. For a set of N face images $x_1, x_2, \cdots, x_N$ with the mean face being μ, the objective is to determine the orthogonal projection ϕ in

$$y_k = \phi^T x_k, \quad k = 1, \cdots, N \tag{2.3}$$

that maximizes the determinant of the total scatter matrix S_T of the projected samples
$y_1, y_2 \cdots, y_N$, where

$$S_T = \sum_{k=1}^{N} (x_k - \mu)((x_k - \mu)^T \tag{2.4}$$

The primary advantage of this technique is that it aids in significant data compression while being sensitive to facial occlusion.

The Bayesian approach proposed by Moghaddam et al. [27], which won the FERET face recognition competition in 1996, uses probabilistic similarity measures by comparing the intrapersonal and extrapersonal variations of face images. The similarity measure $S(I_1, I_2)$ between a pair of images is defined in terms of the intrapersonal a posteriori probability. The approach is robust to expression and illumination variation. However the presence of occlusions increases the dimensionality of the subspaces and degrades the density model which eventually results in misclassification of faces.

Bartlett et al. [31] applied Independent Component Analysis (ICA), an appearance-based technique, for the problem of face recognition. While PCA decorrelates the input data using second-order statistics and thereby generates compressed data with minimum mean-squared reprojection error, ICA minimizes both second-order and higher-order dependencies in the input. It is intimately related to the blind source separation (BSS) problem, where the goal is to decompose an observed signal into a linear combination of unknown independent signals. The objective of ICA is to find the mixing matrix A or the separating matrix W to yield an output vector U using

$$U = Wx = WAs \tag{2.5}$$

where $x = As$ is the mixing model. The sparsity property of ICA basis images makes the performance of ICA better than PCA in terms of robustness to partial occlusions and local distortions, such as changes in facial expression, because spatially local features only influence small parts of facial images [114]. However, ICA basis images do not display perfectly local characteristics in the sense that pixels that do not belong to locally salient feature regions still have some non-zero weight values. These pixel values in non-salient regions appear as noise and contribute to the degradation of the recognition performance specifically when faces are prone to occlusions.

The holistic Linear discriminant Analysis (LDA) which is also known as Fisher Discriminant Analysis (FDA) classifies face images of unknown classes based on training samples with known classes. It aims to maximize between-class variance and minimize within-class variance. Both off-line feature extraction and on-line feature computation can be done at high speeds, and recognition can be done in almost real time using LDA. Mathematically, it calculates the projection matrix W that maximizes the Fisher's Linear Discriminant (FLD) criterion [115] as follows:

$$J_{FLD}(W_{opt}) = \arg\max_{W} \mid W^T S_b W \mid / \mid W^T S_w W \mid \tag{2.6}$$

where S_w and S_b are respectively the within-class scatter matrix and the between-class scatter matrix. When compared to LDA[37], the Efficient Pseudoinverse LDA (EPLDA) [116] is better in handling occlusions as it uses QR decomposition and Discriminant Common Vectors to tackle the singularity problem posed by LDA. But both LDA and EPLDA yields less than 80% recognition rates in the presence of occlusions such as sunglasses or scarf.

Other approaches that mostly capture global features of face images such as Support Vector Machines (SVMs) [117, 118], Kernel Methods [119, 120] and Neural Networks [121, 122] have been used to construct a suitable set of face templates. These approaches suffer recognition performance when faces are prone to occlusions [5, 123] mainly due to the following reasons

- They are characterized by the lack of a-priori decomposition of the image into semantically meaningful components.
- Global features are influenced easily by noise or occlusion.

2.3.2 Component Based Models

Component-based models otherwise known as local models subdivide the object under study into components, then process and manipulate these components, to finally classify them based on one-to-many and many-to-one mappings.

One of the pioneer contributions in recognizing faces with partial occlusions was by Martinez et al. [124, 125]. Firstly they developed a huge dataset with more than 3200 subjects with face images containing real occlusions (Sunglasses and Scarf), which has been used by researchers for experimental validations till now. Secondly the component based Martinez Localization Algorithm (MLA) [125] to recognize partially occluded faces with frontal views serve as a benchmark test [126]. This component based model divides each face into six local regions which are analyzed discretely. A weighted eigenspace representation has been built to overcome expression and occlusion variations. The authors have shown that the method is robust to recognize faces with about one third occlusion (sunglasses or scarf). The method has been better able to handle eye occlusions than mouth occlusions. It demands high computation time due to the use of mixtures of Gaussian distributions. Also the technique relies on manually extracting the ground truth of several facial locations for want of warping the faces.

Kalocsai et al. [39] performed a face recognition experiment in which the performance of a local feature based system, using Gabor-filters, and a global template matching based system, using a combination of PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) was correlated with human performance. Both systems showed qualitative similarities to human performance and the experimental results indicated an important outcome that the preservation of local feature based representation might be necessary to achieve recognition performance similar to that of humans.

As an extension of ICA [31] which has been described in section 2.3.1, Kim et al. [114] have proposed a part-based Locally Salient ICA (LS-ICA) approach to recognize faces with partial occlusions. The authors have stressed that the

"Recognition by parts" paradigm is essential to recognize occluded faces. The kurtosis of U in eq.(2.5) has been defined as

$$kurt(U) = \mid E\{(U)^4\} - 3(E\{(U)^2\})^2 \mid \tag{2.7}$$

The objective of LS-ICA is to maximize the kurtosis defined in eq.(2.7) in order to eliminate non-local modulation imposed on the ICA architecture. The method has been shown robust for recognizing a minor occlusion content of about 10%. The validity of the approach for robustness to a moderate occlusion content is not evident.

Kumar et al. [17] have proposed Correlation Filters (CF) [127, 128] also known as spatial frequency domain methods for face recognition robust to common variations, especially occlusions, as they offer "graceful degradation" owing to the integrative nature of the matching operation they deploy. The cross correlation between a reference pattern $r(x, y)$ and a test pattern $t(x, y)$ for possible shifts τ_x and τ_y has been defined as

$$c(\tau_x, \tau_y) = \int\int t(x, y) r(x - \tau_x, y - \tau_y) dxdy \tag{2.8}$$

where the limits of integration are based on the support of $t(x, y)$. Often, as these two patterns being compared exhibit relative shifts, selecting its maximum as a metric of the similarity between the two patterns yields the discrimination potential for robust pattern recognition tasks such as face recognition. Further Laia et al. [129], motivated by the adaptive beam-forming technique, have proposed component based CFs that can adapt and automatically tune out the actual occlusion (noise/distortion) from test data without making any arbitrary assumptions. CFs have the advantage of yielding a stable correlation peak that changes very little even when there is a large change in the strength of the distortion/noise. If some of the pixels are occluded, they simply do not contribute to the correlation peak, thus decreasing the overall peak. However, no single pixel in the image domain is critical in that recognition can be still carried out successfully. As the filters used are linear, the technique has limitations to learn and classify face images that are not linearly separable. CFs can be computationally complex due to carrying out multiple 2D Fast Fourier Transforms. The class-dependence feature analysis method proposed by Kumar et al. [17] attempts to mitigate this complexity to a certain extent. CFs are not suitable to apply in situations where the training data is sparse. A further drawback is that CFs do not take advantage of domain-specific knowledge about face images. Hence future research should focus on fusing image-domain approaches with feature-based approaches.

The initial occlusion model proposed by Zhang et al. [130] using Local Gabor Binary Patterns (LGBP) failed to recognize faces with upper occlusions as the occluded probability of local regions were not taken into account. As a remedy the Kullback-Leiber Divergence (KLD) based LGBP approach [131] has been recently proposed to estimate the probability of occlusion and uses weights of local regions for the final feature matching process. However the main drawback of the approach is its high dimensionality.

Hotta et al. [123] have presented an approach using SVM with Local Gaussian Summation Kernel (SVM-LGSK) for recognizing faces with partial occlusion. The SVM determines the optimal hyperplane which maximizes the margin, where the margin is the distance between the hyperplane and its nearest sample. For the training set and its label denoted as $S = \{(x_i, y_i), \cdots, (x_L, y_L)\}$, the optimal hyper plane is defined as

$$f(x) = \sum_{i \in SV} \alpha_i y_i x_i^T x + b, \tag{2.9}$$

where SV is a set of support vectors, b is a threshold and the non-zero support vectors are represented by α. In the proposed method, local kernels are arranged at all positions on the face. Each local kernel plays the role of visual cells specialized for local features of each person's face. In order to develop the visual cells specialized for local features, the stimulus selectivity of a Gaussian kernel is suitable. Hence, a Gaussian kernel is used as the local kernel. The local Gaussian kernel is defined by

$$K_p((x(p), y(p)) = exp\left(\frac{- \parallel x(p) - y(p) \parallel^2}{2\sigma_p^2}\right) \tag{2.10}$$

where $x(p)$ and $y(p)$ are the local features centered at label of position p and σ_p^2 is the local variance at p. The optimal hyperplane of SVM with local Gaussian summation kernel is defined by

$$f(x) = \sum_{i \in SV} \alpha_i y_i \frac{1}{N} \sum_{p}^{N} exp\left(\frac{- \parallel x(p) - y(p) \parallel^2}{2\sigma_p^2}\right) + b, \tag{2.11}$$

where N is the number of local kernels. With this formulation, keeping SVM as a binary classifier, face recognition is performed. Unlike the global kernel, the local Gaussian summation kernel is not influenced by noise or occlusion and hence the approach is feasible for occluded face recognition. The selection of an appropriate size of the local kernel depends on the position and recognition target which seems to be a bottle-neck in the proposed method.

A Selective Local Non-negative Matrix Factorization (S-LNMF) technique has been proposed by Oh et al. [132] to attack the occlusion problem. The basic idea is that local occlusion affects only the coefficients of the corresponding local bases and hence the error caused by occlusions are local and not global in nature. By using the LNMF bases for occlusion-free regions exclusively, occlusions have been detected. Each face image,Ω, has been divided into six local disjoint patches and their PCA coefficients are computed by

$$\Omega_{i,k} = E_k^T (X_{i,k} - \psi_k), \qquad i = 1, 2, \cdots N, \quad k = 1, 2, \cdots 6, \tag{2.12}$$

where $X_{i,k}$ is the k^{th} patch of the i^{th} image, ψ_k, E_k are the mean image and the eigen-matrix of the k^{th} patch and N is the total number of training images. The occlusion detection for each patch is accomplished by comparing the coefficient vectors of occlusion-free images with that of the test image in the corresponding eigenspace. To detect the bases in the occluded regions, an occluded energy measure per basis image has been defined as

$$E^{i}_{Occlusion} = \frac{\sum_{x,y\in W} I_i^2(x,y)}{\sum_{x=1}^{C}\sum_{y=1}^{R} I_i^2(x,y)}, \quad i = 1, 2, \cdots, N \tag{2.13}$$

where $C \times R$ is the image size, $I_i(x,y)$ is the value of the i^{th} LNMF basis at (x,y), W is the detected occluded region. The occluded energy value serve as a clue to know whether an LNMF basis image has been occluded or not. By excluding the coefficients of occluded parts, the effect of occlusion in the final match is minimized. The approach shows improved recognition rates over many standard techniques including LFA[133], R-PCA [134] and LNMF[135]. The main drawback of the approach is that the computation of occluded energy eq.(2.13) demands huge computing time (100 times more than PCA).

The algorithm proposed by Wright et al. [136] exploits the fact that errors caused by occlusion typically corrupt only a fraction of the image pixels and hence yield a sparse representation which aids in better manipulation of occlusions. The linear representation of a probe face y has been represented as $y = Ax_o + e_o$ where A is a matrix which linearly spans the training samples, x_o is a coefficient vector and e_o is a vector of errors. It has been assumed that e_o can be sparsely represented as $e_o = A_e u_o$ where u_o is some sparse vector. While this Sparse Representation based Classification (SRC) approach attacks the occlusion problem well, it has a drawback to handle pose variations as the number of training samples to represent the pose variation can be prohibitively large. In the future, one may adopt the Active Appearance Model (AAM) [137] which is capable of tracking face images which are prone to occlusion as well as pose variations for the occluded face recognition problem.

Kanan et al. [126] have proposed a component model based on Adaptively Weighted Sub-Gabor Array (AWSGA) when only one sample image per enrolled subject is available. The proposed algorithm utilizes a local Gabor array to represent faces partitioned into sub-patterns. For a given face $f(x,y)$ of dimensions $N \times N$ at orientation θ and radial center frequency ω_0 which is segmented into sub-patterns of height W, its SGW representation has been defined as

$$SGW^{P}_{\omega_o,\theta}(x,b) = F^{-1}\{F\{f(x, W(p-1)+b)\}F\{\psi_{\omega_0,\theta}(x,y)\}\} \tag{2.14}$$

where $S = N/W$, $1 \leq p \leq S$, F, F^{-1} are Fourier and inverse Fourier transforms respectively and ψ is a 2D Gabor Wavelet which is defined as

$$\psi_{\omega_0,\theta}(x,y) = S_{\omega_o,\theta}(x,y)W_{\omega_0,\theta}(x,y) \tag{2.15}$$

The general idea behind the proposed method relies on the observation that occlusions appear as local distortions away from a general face representing human population. This distortion measurement is utilized in the proposed approach for weighting individual Sub-Gabor elements. A Sub-Gabor Wavelet (SGW) operation is performed on a partitioned image to form an Augmented SG Array (ASGA) of the face image. While the approach can slightly improve lower face occlusions, still it suffers from upper face occlusions.

The above findings show the ”divide and rule” phenomenon adopted by component based face recognition models. In essence, component based approaches have two main common features. Firstly they can efficiently represent specific components of an object as discrete entities. Secondly they have sound integrating mechanisms to relate these components to determine the object class. This unique way of modularly processing objects leads to recognizing objects subject to uncertainties including occlusions.

2.3.3 What is the State of Present Occlusion Models?

The dataset developed by Martinez [124] has been used by number of researchers especially to validate an algorithm against two real occlusions namely sunglasses and scarf. We have compared the performance of the following state-of-the-art face recognition models with respect to this standard occlusion test and chronologically projected them in Fig.2.1:

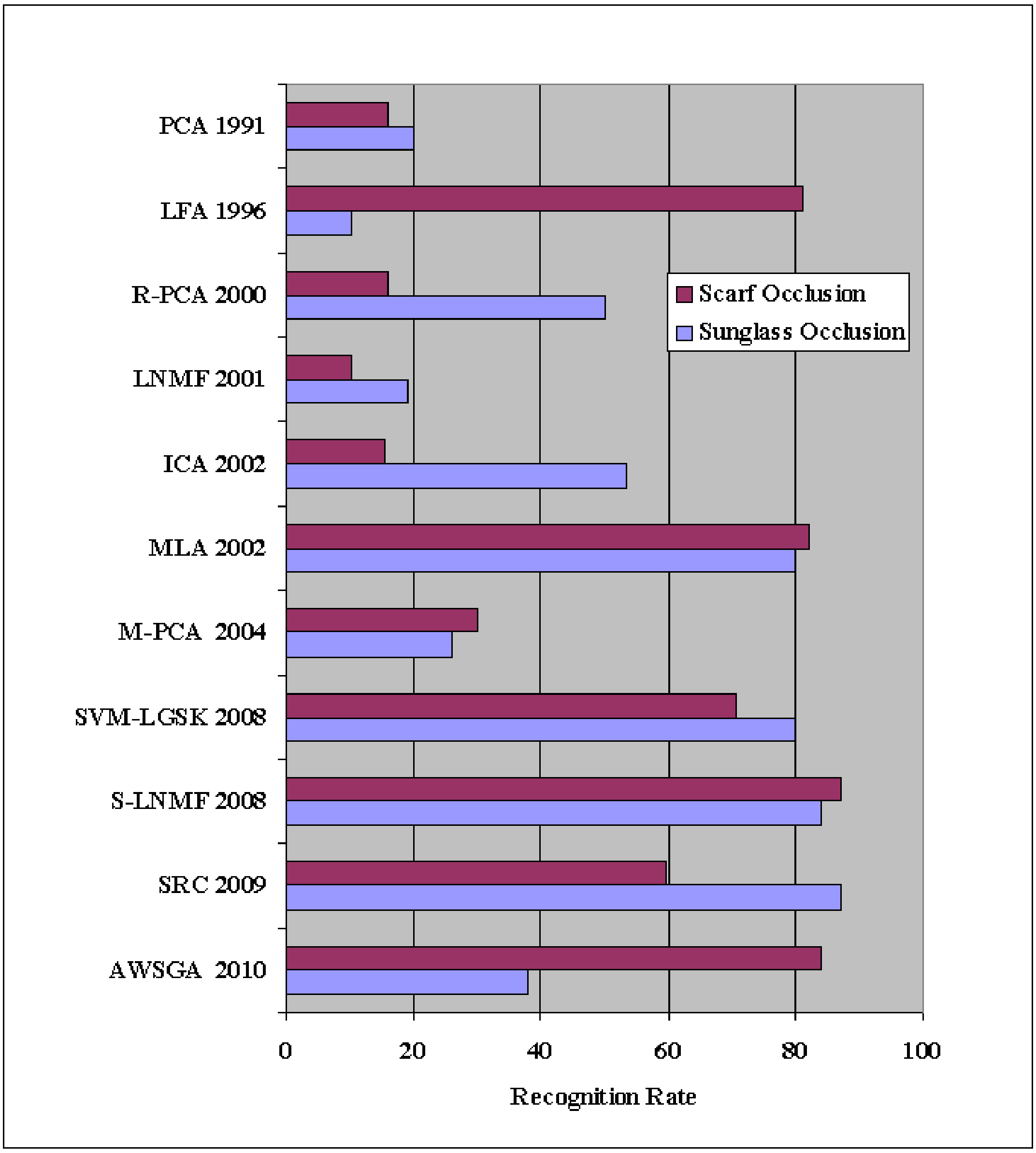


Figure 2.1: Comparison of State-of-the-art Occlusion Models using the AR Face Dataset

i) Adaptively Weighted Sub-Gabor Array (AWSGA) [126]

ii) Sparse Representation based Classification (SRC) [136]

iii) Selective Local Non-negative Matrix Factorization (S-LNMF) [132]

iv) Support Vector Machine with Local Gaussian Summation Kernel (SVM-LGSK) [123]

v) Modular Principal Component Analysis (M-PCA) [138]

vi) Martinez Localization Algorithm (MLA) [125]

vii) Independent Component Analsysis (ICA) [31]

viii) Local Non-negative Matrix Factorization (LNMF) [135]

ix) Robust Principal Component Analysis (R-PCA) [134]

x) Local Feature Analysis (LFA) [133]

xi) Principal Component Analysis (PCA) [26]

The chart provides a bird's eye view of how various algorithms can tolerate occlusions. Further the chart shows the fact that significant improvement has been seen as a result of evolving occlusion models since the commencement of the first automated face recognition system in 1991 [26] till today. However even under control conditions, for a data set of around 3000 face images, the current state-art-of-the-art cannot yet yield promising recognition rates when face images are prone to occlusions such as sunglasses or scarf. We see that algorithms that are good in handling lower face occlusions need not have to be good in handling upper face occlusions and vice versa.

2.4 Summary

We have seen that psychologically feasible computational models exhibit clear and strong relationships between behavior and properties of the domains which they intend to represent. Psychological experiments reveal that humans have the ability to recognize complex occluded objects by processing them in terms of subregions. This object-based mechanism has been adapted by many computer vision techniques that address uncertainty issues such as occlusion. Psychophysical studies indicate that humans as well as machines benefit by processing faces in terms of horizontal orientation. Some of the popular imaging applications that use face processing algorithms are based on the principles of cognitive psychology. A widespread adoption of probabilistic models in many areas of computer vision and pattern recognition has been witnessed over the last decade. Bayesian graphical models have been applied by a number of computer vision applications that work in unconstrained and realistic environments and they are now the mainstay of the AI research field known as “uncertain reasoning”.

Our literature review reveals that, technically, holistic methods intend to classify objects by relying on some linear or nonlinear transformations on the holistic image vectors used for training and are shown to be robust against global variations such as lighting or aging effect. However, they may not fit well with images with partial occlusions because of the fact that the resulting holistic representations are usually deviated far from the normal patterns. Hence, to handle occlusions and other intricacies of real world scenarios such as noise and cluttered background, research focus has shifted from holistic processing to component based representations. Literature shows that component based approaches provide robust means to counter the occlusion problem by efficient representation, reasoning and intelligent decision making mechanisms under uncertainty.

Further, we have systematically synthesized, analyzed and presented an up-to-date overview of various occlusion models and clearly demonstrated where they stand after being progressively evolved and improved over the past few years. Notwithstanding the progressive research effort that has gone into the modeling of occluded face recognition algorithms, we are yet to see a system that can be deployed effectively to recognize faces which are prone to multiple occlusions in an unconstrained setting. Still, recognizing faces with partial occlusions remains a partially solved problem. Lessons learnt from related work, as an outcome of this chapter, motivate and spearhead us to formulate a robust face recognition model, the details which we will present in Chapter 3.

Chapter 3

PROCIM for Robust Face Recognition

3.1 Introduction

In this chapter, we will provide the theoretical constructs of the proposed **PRO**babilistic **C**omponent **I**nterpretation **M**odel (PROCIM). In chapter 2 we have identified some of the core principles from cognitive psychology and neuroscience discipline with relevance to reasoning under uncertainty. Based on these principles we will initially present a conceptual human model to understand how humans recognize complicated objects and then gradually transform this model into a psychologically plausible probabilistic model using Bayesian Networks.

3.2 How do Humans Recognize Complex Objects?

Though a major portion of a face is occluded, a human being, by evidencing small subregions of the face, could still recognize the face. This remarkable recognition ability is governed by the following key principles which we recall from Section 2.1.

* Humans, in order to minimize the processing load, organize a complex occluded object into subregions and then attend selectively to particular physical regions. This selective attentional spotlight focuses on areas of interest and facilitates preferential processing of information from those chosen areas [40].

* Humans perform substantially better at identifying faces that have been filtered to contain just horizontal information compared to any other orientation band and processing faces in terms of horizontal structures has computational advantages [54].

* Similarity is a basic concept in cognitive psychology which is utilized to explore the principles of human perception [65]. Recent studies [66, 67] that refer to the classical contrast model of similarity [68] insist that perceived similarity is the result of a feature-matching process. A fundamental hypothesis associated with the perception and memory of faces states that, humans perceive and remember faces by means of facial features [69].

* Facial discrimination tasks performed by the human visual cortex rely on the combination of multiple local feature analyzers rather than global information [64].

* *"Increasing familiarity"* is an important factor which enables humans to recognize highly degraded face images [28].

The above principles reveal the fact that humans gaze at ambiguous faces in stages to gather cues and map them with the facial domain (numerous faces they remember) to recall similar faces. In this way they could recall (shortlist) a subset of faces which closely resembles the features of these subregions, out of the huge number of known faces. With prior beliefs about faces assimilated from experience, humans finally rank-list the most probable faces from the shortlisted faces with preferences. An example of this scenario is depicted in Fig. 3.1. When a human observer encounters a face with major occlusions like the one shown in Fig. 3.1,

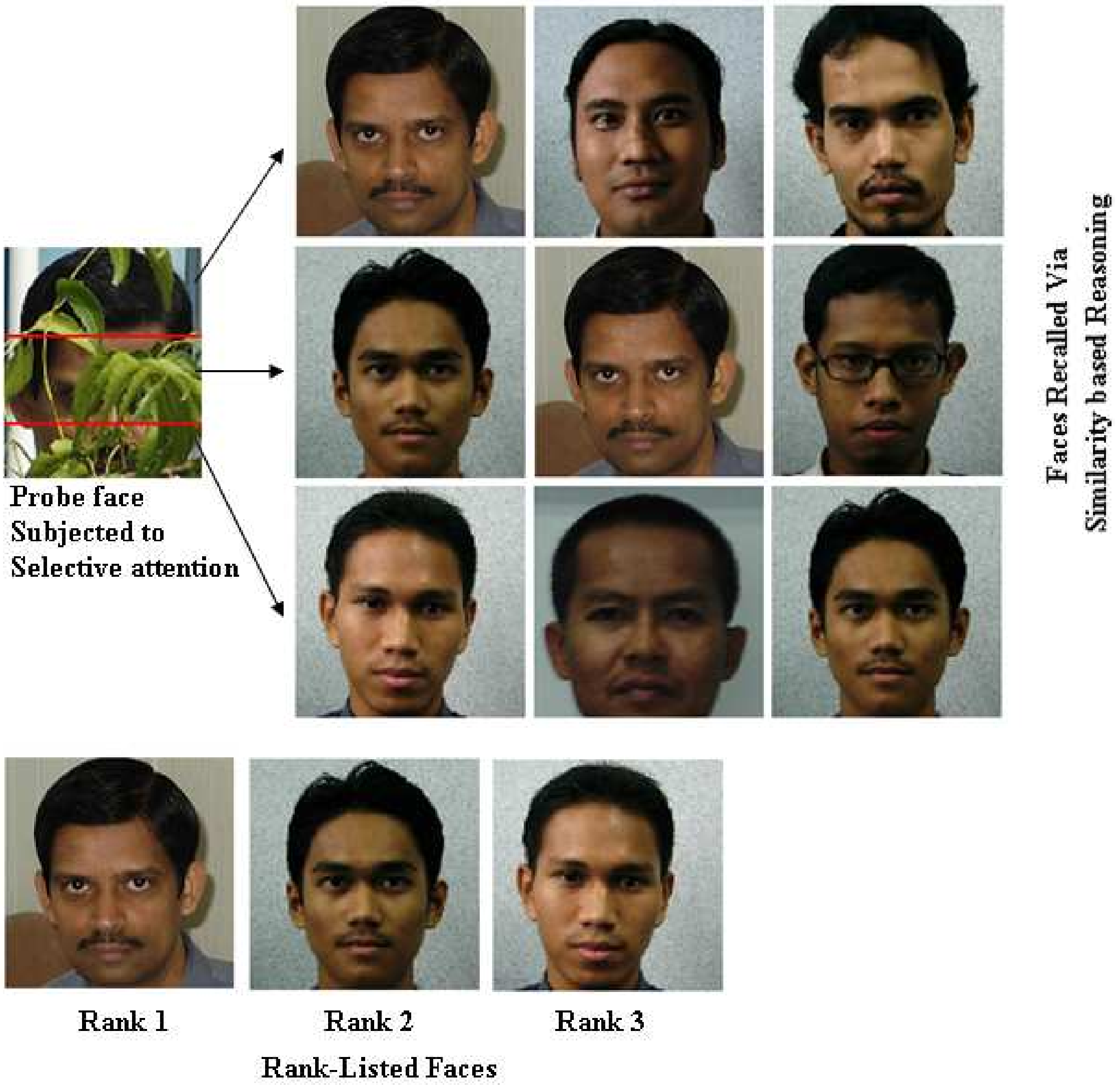


FIGURE 3.1: A typical scenario depicting human reasoning based on similarity mapping; A human observer gaze the probe face and recalls few faces. Uncertainty factors such as occlusion can muddle up the order(ranks) in which faces are recalled. It is not necessary that every gaze should recall the correct face in the first instance (rank). Finally the human might make a decision amidst uncertainty by analysing influence strengths

where leaves of a tree occludes the probe face, uncertainty arises as few cues of the facial features are visible. For example, let the probe face be gazed (observed) at three horizontal subregions (top, mid and bottom as separated by red lines in Fig. 3.1) by the human observer. The human observer might counter the uncertainty by establishing and exploiting similarity maps between the subregions of the occluded face and the huge number of faces, which is the face domain, he or she knows. These similarity mappings in turn influence a restricted subset of shortlisted faces and the degree of uncertainty gets considerably reduced. Referring to the recalled (shortlisted) faces in Fig. 3.1, the actual suspect's face is influenced by the top most subsample in the first rank because of the fact that more visual cues are present in the subsample. More occlusions leads to more ambiguities in the cues, as shown by the mid subsample. Consequently the actual suspect's face has been influenced in the second rank by the mid subsample. As the severity of occlusions increase the actual suspect is not influenced at all by the bottom most subsample, which is the worst case. In the context of face recognition, the actual suspect need not always be in the first rank [2]. As long as the actual suspect is within an acceptable range, a solution is still arrived at. Finally a decision is made by probabilistic means by manipulating the influence strengths exhibited by the shortlisted faces and a few are rank listed as being recognized. We will scientifically define influence strengths in the following section (3.3). The study of such human reasoning reveals an important hypothesis. That is by mapping intrinsic similarities between the set of subsamples of the probe face and the set of faces in the facial domain and analysing the influence strengths, humans might be able to recognize faces with major occlusions reasonably well.

3.3 Framework of the Proposed PROCIM

Basically PROCIM intends to map similarities between two main object entities. The first entity is a set of sparse components of the probe face image. The second entity comprises a bulk set of known face images that are stored in the face database which are known as gallery samples. Few faces are recalled for each of the sparse components as a consequence of this similarity based reasoning. Further this reasoning reveals inherent conditional independence properties between the recalled face images. These independencies could be scientifically represented by Bayesian Networks (BNs) as mentioned by Nilsson [139]. Formally a BN is a Directed Acyclic Graph(DAG) where the nodes represent variables and the arcs encode conditional independencies between the variables. BNs serve as fundamental tools in tackling uncertainty problems as they characterize intuitive notions of human reasoning. In other words, PROCIM employs BNs to establish and learn intrinsic similarity mappings that are inherent in the face domain. Then PROCIM takes robust probabilistic decisions by exploiting the mappings established.

We will briefly present the framework of the proposed PROCIM with the aid of the flow-chart shown in Fig.3.2. Firstly, face images in the database have been enhanced with standard preprocessing techniques. We have adapted the techniques proposed by [31] to preprocess face images. Standard datasets provide ground truth data of eye and mouth coordinates. Basically these coordinates were used to center and align the face images, and then crop and scale them to standard dimensions without the need of manual intervention. This image preprocessing enables the face images to be independent of variations such as scaling, translation, rotation and so on. Then the feature space (low dimensional face space) is constructed from the gallery (training set) of face images available in the face database using Principal Component Analysis (PCA). PROCIM further learns conditional probability potentials which provide information about how well a face can be influenced given that a particular face component or a combination of face components has been observed. As the DAG of a BN is called the structure and the values in the conditional probability distributions are called the parameters, learning the conditional probability potentials otherwise means parameter estimation. This learning process is done offline from the gallery face images, that is when computing resources are free.

Each node of a BN has a set of probable values for each variable which are known as belief states. These belief states are propagated between nodes of the BN effectively. A BN is good at mapping intrinsic relationships that are inherent in a domain in terms of parent and child nodes. The intuitive meaning of an arrow (arc) from a parent node to a child node(s) indicates that the parent node has influenced the child node(s)[140]. The learned belief states are stored in Conditional Probability Tables (CPTs). Thus PROCIM is capable of learning prior information and

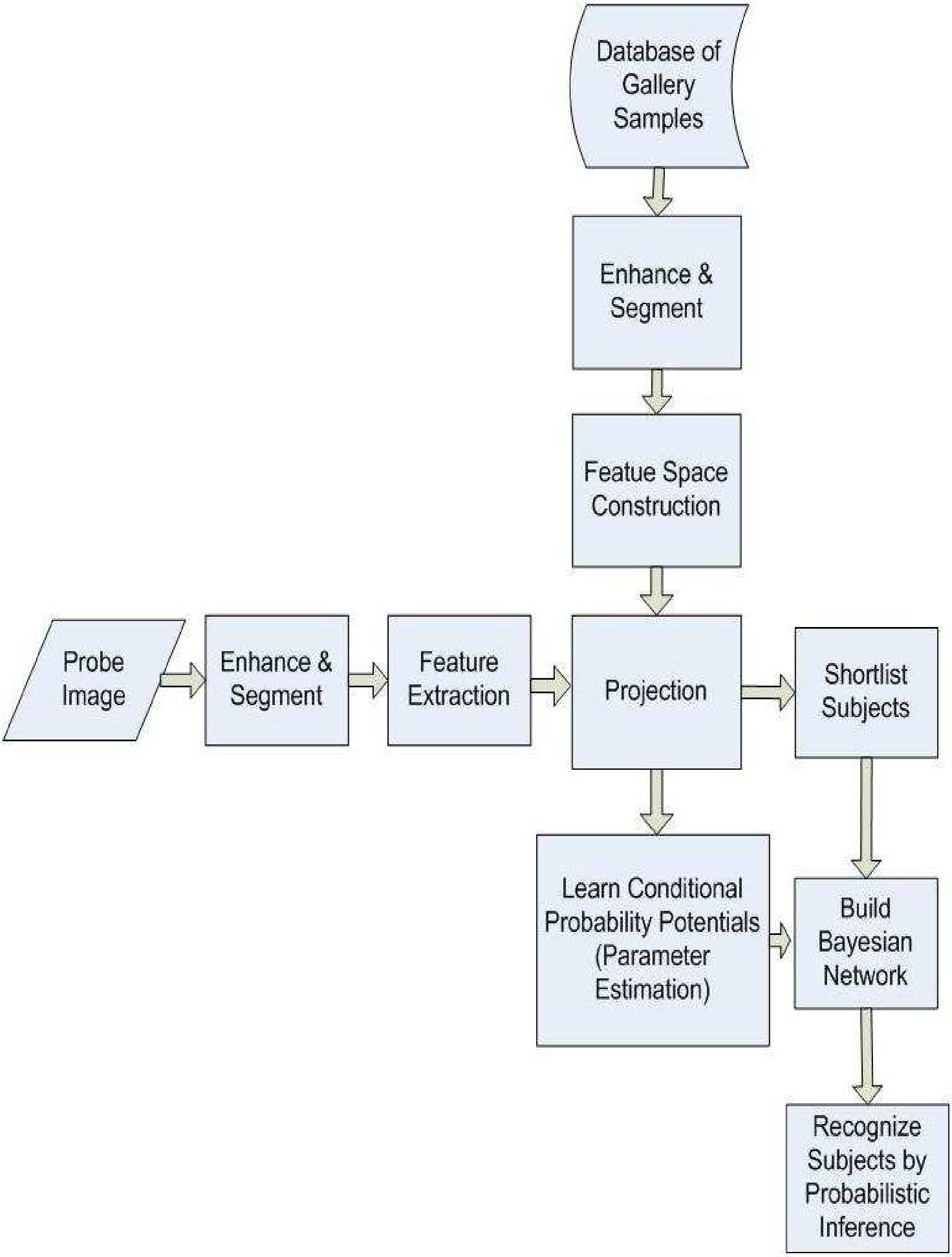

Database of Gallery Samples
Enhance & Segment
Featue Space Construction
Probe Image
Enhance & Segment
Feature Extraction
Projection
Shortlist Subjects
Learn Conditional Probability Potentials (Parameter Estimation)
Build Bayesian Network
Recognize Subjects by Probabilistic Inference

Figure 3.2: Flowchart showing the various stages of the proposed PROCIM architecture

experience about facial domains. A given probe face is enhanced using similar preprocessing techniques which were applied to the gallery samples and subject to horizontal segmentation. Then the PCA features of facial entities, acquired by combining probe face components over the gallery face images, are extracted and projected over the feature space using an inheritance mechanism which will be described in Section 3.5.1. Further probable subjects are shortlisted by means of similarity mapping based processing. For a given probe, a BN is generated whose child node variables represent the belief states of short-listed subjects and the parent nodes represent the belief states of corresponding components which influenced them. Finally faces are rank-listed using a face score formula which will be derived in Section 3.7.

3.4 Building the Bayesian Network from a Simple Component Based Scheme

We have portrayed systematically how humans might exploit the similarity mappings that exists naturally in the face domain to recognize faces with major occlusions in Section 3.2. In this section we will present the feasibility of application of BNs to counter uncertainties and gradually show how they aid in scientifically modeling the intuitive similarity mappings. The application of BNs to uncertainty problems offers the following advantages:

i. They provide a simple way to visualize the structure of an abstract probabilistic model and can be used to design and motivate new models. The benefit lies in the way such a structure can be used as a compact representation for many naturally occurring complex domain problems, specifically recognizing occluded faces and gaits in unconstrained environments.

ii. Insights into the properties of the model, including conditional independence properties, can be obtained by inspection of the graph. We will shortly show in Section 3.8 that the BN generated by PROCIM aids us to visualize and analyze the impact of occlusion and other variations coherently.

iii. Complex computations, required to perform inference and learning in sophisticated models, can be expressed in terms of graphical manipulations, in which underlying mathematical expressions are carried along implicitly.

iv. It is intuitively easier for a human to understand the network structures and the local distributions via BNs than complete joint distributions. Further BN structures provide the flexibility to modify them, if necessary, in order to obtain better predictive models.

v By adding utility functions, the BN model can be extended to decision networks for decision analysis. We will shortly show in Section 3.7, how PROCIM can take a meaningful decision amidst uncertainty.

An important concept for probability distributions over multiple variables is that of *conditional independence* [141]. Conditional independence properties play an important role in using probabilistic models for pattern recognition by simplifying both the structure of a model and the computations needed to perform inference and learning under that model. An important and elegant feature of graphical models is that conditional independence properties of the joint distribution can be read directly from the graph without having to perform any analytical manipulations. By manipulating the belief states in the BN, the state of a particular node can be queried from other nodes with the aid of probabilistic inference techniques. In our case we would like to query the belief state of an occluded probe face or a gait subjected to complexities by observing the probabilities entailed by its subsamples (components).

Diverse sources of information content are possessed in various subregions of a face. Owing to this variation, though we segment the face into equal horizontal rectangular subregions, not all these subregions will have the same probability of influencing the face to be recognized([142–145] as referred in [28]). These references show that the order of influence strengths ranges from eyes, followed by mouth and then the nose. Therefore each subsample (in statistical sense a subsample refers to a subregion) of the face will have different belief states. The more unique features a subsample might contain, the more strength it might be have to influence the face. This is the key reason why human beings, by just seeing a small portion of a face, while other salient portions of the face might be occluded, can recognize faces with major occlusions. We define the strength of a subsample which crucially contributes in influencing the recognition mechanism of the face as *Influence Strength* and denote it as Z. We will define subscripts of Z later. How well a subsample can influence the face which encodes it depends upon the physical properties of the subsample. The definition of Z here strongly conveys the physical phenomenon associated with the recognition mechanism. This leads us to hypothesize that *"The face being recognized by observing a subsample of an occluded probe face will be more similar to the probe, if Z's magnitude is high"*.

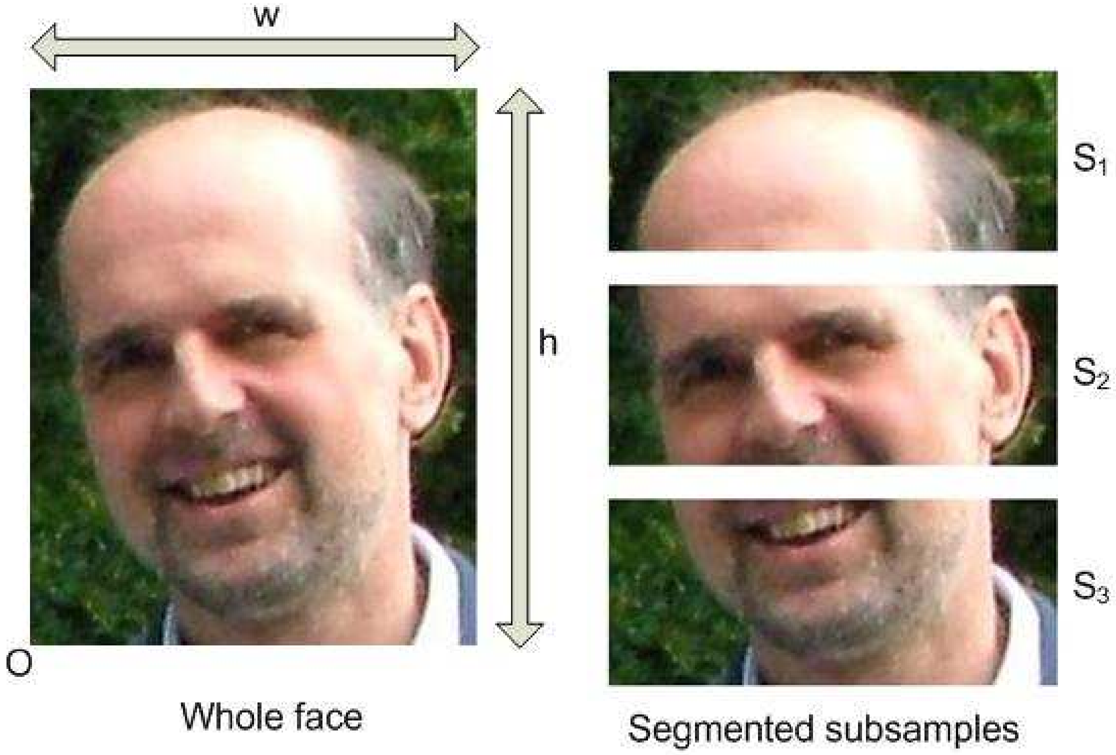


FIGURE 3.3: The proposed psychologically plausible segmentation scheme

We have seen in Section 2.1 that cognitive and neuroscience literature reveals certain principles such as selective attention, advantages of processing faces in terms of horizontal structures and human visual cortex's reliance on the combination of multiple local features. These facts have been further reinforced in Section 3.2 which portrays human strategies about reasoning under uncertainty. Further in Section 2.3.1 we have seen that global features are influenced easily by noise or occlusion as they are characterized by the lack of a-priori decomposition of the image into semantically meaningful components. The above findings show the "divide and rule" phenomenon adopted by component based face recognition models. After a thorough review of component based models in section 2.3.2 it is evident that these local approaches can efficiently represent specific components of an object as discrete entities and they have sound integrating mechanisms to relate these components to determine the object class. This unique way of modularly processing objects leads to recognizing objects subject to uncertainties including occlusions. We consolidate all these findings gradually into PROCIM's architecture as follows.

Let the probe face be segmented into k equal horizontal rectangular subregions. Let the k subsamples of the probe face be represented by $S = \{S_1, S_2, S_3, \cdots, S_k\}$. A typical face subjected to horizontal segmentation for the case of $k = 3$ is shown in Fig. 3.3. Let $F = \{F_1, F_2, F_3, \cdots, F_n\}$ represent the training face set which has face images of n subjects. Suppose that a subsample $S_i \subset S$, $1 \leq i \leq k$, has influenced the recognition of a set of faces $f = \{F_p, F_q, F_r\} \subset F$, where p, q and r represent unique integers between 1 and n. Let Z_{ip}, Z_{iq} and Z_{ir} represent the corresponding influence strengths as shown in Fig. 3.4. We refer to Fig.3.3 where a typical face image, say $I(x, y)$ has been segmented into three subsamples. Let $O(x_0, y_0), h, w$ respectively represent the co-ordinates of the bottom left point, height and width of $I(x, y)$. We will use the set theory notation [146] which is widely referred to in the literature to define the subsamples $S_1, S_2, S_3, \cdots, S_k$ as follows:

$$S_1(I(x,y)) = \left\{(x,y) | x_0 \leq x \leq x_0 + w, (y_0 + h) - \tfrac{h}{k} \leq y < y_0 + h\right\}$$

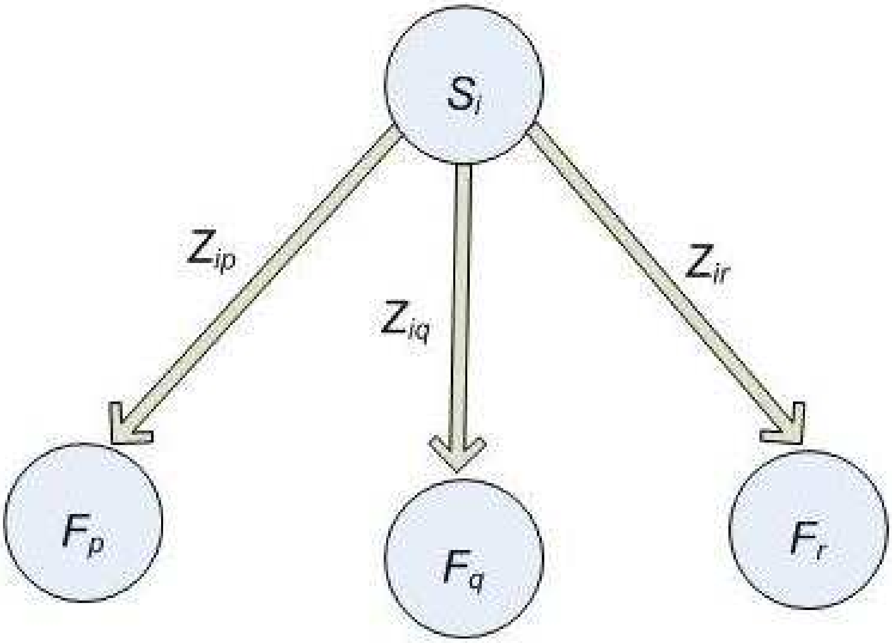


FIGURE 3.4: Proposed DAG model showing mappings between a subsample and the faces being recognized as a consequence of its influence; This Markov conditioned DAG is nothing but a BN.

$$S_2(I(x,y)) = \left\{(x,y)|x_0 \le x \le x_0 + w, (y_0+h) - \tfrac{2h}{k} \le y < (y_0+h) - \tfrac{h}{k}\right\}$$

$$\vdots \tag{3.1}$$

$$S_k(I(x,y)) = \left\{(x,y)|x_0 \le x \le x_0 + w, y_0 \le y < (y_0+h) - \tfrac{(k-1)h}{k}\right\}$$

Interestingly a basic understanding of graph theory fundamentals will reveal that the graph shown in Fig. 3.4 will constitute a Directed Acyclic Graph (DAG) [147]. The pair (S, E) constitutes a directed graph, where S is a finite, nonempty set whose elements are called nodes (or vertices), and E is a set of ordered pairs of distinct elements of S. For the graph shown in Fig. 3.4, $E = \{(S_i, F_p), (S_i, F_q), (S_i, F_r)\}$, where elements of E are called edges (or arcs). $\forall(X,Y) \in E, X$ and Y are each incident to the edge (X,Y). Suppose we have a set of nodes $[X_1, X_2, \cdots, X_k]$, where $k \geq 2$, such that $(X_{i-1}, X_i) \in E$ for $2 \leq i \leq k$. We call the set of edges connecting the k nodes a path from X_1 to X_k. The nodes $X_2, \cdots, X_{k-1}$ are called interior nodes on path $[X_1, X_2, \cdots, X_k]$. The subpath of path $[X_1, X_2, \cdots, X_k]$ from X_i to X_j is the path $[X_i, X_{i+1}, \cdots, X_j]$, where $1 \leq i \leq j \leq k$. A directed cycle is a path from a node to itself. A simple path is a path containing no subpaths which are directed cycles. A directed graph G is called a DAG if it contains no directed cycles. Given a DAG $G = (V, E)$ and nodes X and Y in V, Y is called a parent of X if there is an edge from Y to X. However, Y is called a descendent of X and X is called an ancestor of Y if there is a path from X to Y. Y is called a nondescendent of X if Y is not a descendent of X.

In Fig. 3.4, since S_i is influencing the recognition of f, we draw edges from S_i to the elements of f, to form the graph shown in Fig.3.4(b). The above definitions (from graph theory), clearly justifies that this graph is a DAG. This DAG helps us to establish mappings from the set of subsamples (S_i) to the subset of faces activated (f). Conceptually these faces will be nearly similar to the probe face which represents these subsamples and hence we call these mappings as similarity mappings. In the DAG shown in Fig.3.4 each face is conditionally independent of the other faces given its parent. That is $I_P(\{F_p\}, \{F_q, F_r\}|S_i), I_p(\{F_q\}, \{F_r, F_p\}|S_i)$ and $I_p(\{F_r\}, \{F_p, F_q\}|S_i)$, where we denote independence of random variables by I_P. This can be precisely written in the following general form

$$P(F_j|F_c, S_i) = P(F_j|S_i), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n. \tag{3.2}$$

where $F_c = F \setminus F_j$. Let the DAG shown in Fig. 3.4 be named as D and its underlying probability distribution be named as P. Then (D, P) satisfies the Markov condition provided by (3.2), as each element of D is conditionally independent of the set of all its nondescendents given the set of parents. Such a Markov conditioned DAG leads to what is known as a Bayesian Network by definition [147]. The graphical nature of PROCIM model helps us to visualize the abstract intrinsic similarity relationships that exists in a facial domain, as a consequence of mapping a subsample S_i to the set of faces f.

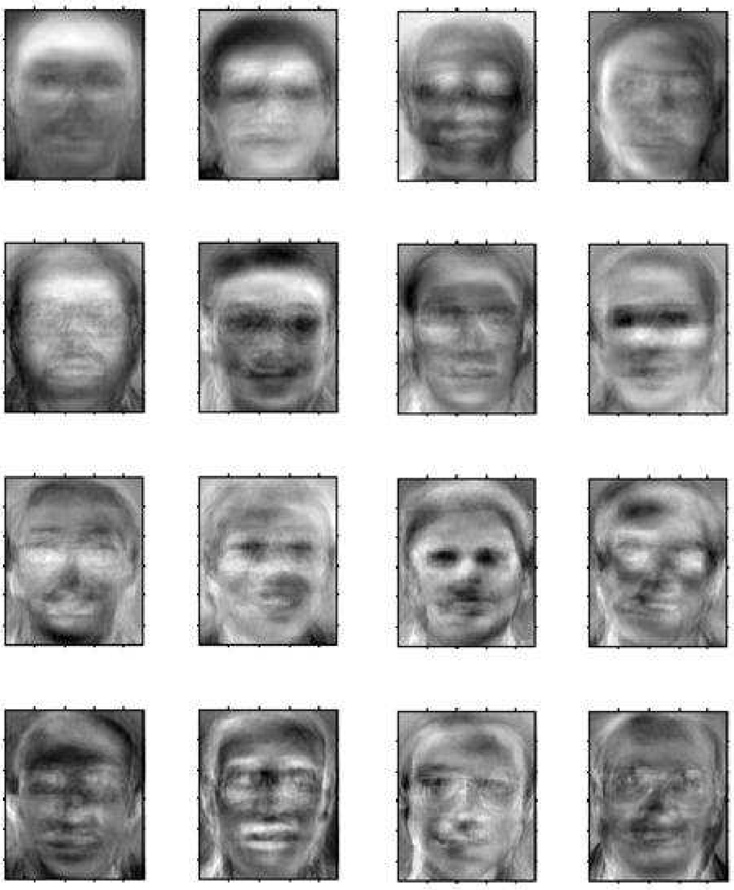


FIGURE 3.5: Ghost like eigenfaces of some typical face images; Technically these are the principal components extracted from face images by applying PCA

3.5 Principal Component Analysis (PCA) based Feature Space

The proposed PROCIM model can be fitted into any suitable feature space projection technique (eg. PCA, ICA, LDA and so on). For example's sake we have chosen the well known PCA architecture. As PROCIM will inherit the PCA architecture for its component based face processing, it is fundamental to describe the basics of PCA here within the scope of the Face Recognition (FR) problem. PCA or Karhunen-Loeve transformation [148] is a standard technique used in statistical pattern recognition and image processing for data reduction and feature extraction [149]. As the input pattern often contains redundant information, mapping it to a feature vector can get rid of this redundancy and yet preserve most of the intrinsic information content of the pattern. These extracted features have a great role in distinguishing input patterns. Hence PCA has been employed as a core technique by most of the successful face recognition techniques such as eigenfaces [26], holons [150] and local feature analysis [133]. PCA based feature space method [151, 152] which is also called as eigenface technique [26, 153] is an appearance-based technique widely used for the dimensionality reduction which has shown a great performance in face recognition.

A face image in two dimensions with size $N \times N$ can also be considered as one dimensional vector of dimension N^2. For example, a typical face image with a resolution of 112 × 92 can be considered as a vector of dimension 10,304, or equivalently a point in a 10,304 dimensional space. An ensemble of images maps to a collection of points in this huge space. Images of faces, being similar in overall configuration, will not be randomly distributed in this huge image space and thus can be described by a relatively low dimensional subspace. The main idea of the PCA is to find the vectors that best account for the distribution of face images within the entire image space. These vectors which define the feature space of face images, is also called face space in FR terminology. Each of these vectors, is a linear combination of the original face images. Because these vectors are the eigenvectors (principal components) of the covariance matrix corresponding to the original face images, and because they are face-like or ghost-like in appearance, they are called eigenfaces [26] though they do not necessarily correspond to features such as eyes, ears and noses. Typical eigenfaces generated from the AT & T face dataset (more details will be provided in Section 3.9.1.1) are shown in Fig. 3.5. Recognition is performed by projecting a new image into the subspace spanned by the feature space. Each face can be approximated using only the eigenfaces which have the largest eigenvalues, that is the best eigenfaces, and therefore account for the most variance within the set of face images. The best M eigenfaces span an M-dimensional feature space of all possible face images.

The mathematical formulation of the PCA approach is as follows: If the gallery set of face images is represented by $\Gamma_1, \Gamma_2, \Gamma_3, \cdots, \Gamma_M$, then their mean face Ψ can be computed using

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \tag{3.3}$$

Each face differs from the average by the vector

$$\Phi_i = \Gamma_i - \Psi \tag{3.4}$$

This set of huge vectors is then subject to PCA, which seeks a set of M orthonormal vectors, u_n, which best describes the distribution of the data. The kth vector, u_k, is chosen such that

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \Phi_n)^2 \tag{3.5}$$

is a maximum. The vectors u_k and scalars λ_k are the eigenvectors and eigenvalues, respectively, of the covariance matrix C which is given by

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T \tag{3.6}$$

$$= AA^T \tag{3.7}$$

where the matrix $A = [\Phi_1 \Phi_2 \cdots \Phi_M]$. The matrix C, however is $N^2 \times N^2$, and determining the N^2 eigenvectors and eigenvalues is a computationally expensive task. But the practical applicability of eigenfaces stems from the possibility to compute the eigenvectors of C using an efficient strategy proposed by Turk and Pentland [26], which is as follows. The rank of the covariance matrix is limited by the number of training examples: if there are M training examples, there will be at most $M - 1$ meaningful eigenvectors with non-zero eigenvalues. By solving for the eigenvectors of an M by M matrix and then taking appropriate linear combinations of the face images Φ_i, the computations are greatly reduced. Thus the associated eigenvalues allow us to rank the eigenvectors according to their significance in characterizing the variation among the images.

3.5.1 Inheriting Similarity Mappings from the PCA based Feature Space

Here we will show how the above phenomenological similarity mapping concept can be brought into reality in a facial (pixel) domain. As the eigenspace is built with the eigenfaces, we cannot directly project the subsamples S_i which do not represent the whole face into this feature space. Building feature spaces for each of the samples is a tedious and roundabout process as well. Rather we strategically combine the subsamples into each of the faces in the FDB and project this combined face, say X_{ij}, onto the eigenspace. Fig. 3.6 shows how a subsample of a typical occluded face has been combined with faces in the training face set. Hence X_{ij} is given by

$$X_{ij} = S_i \cup F_j, \quad i = 1, \dots, k, \quad j = 1, \dots, n. \tag{3.8}$$

Finally a set of r similar faces from the FDB of n face images are shortlisted. That is we intend to shortlist r similar faces (closely resembling the probe face) from the FDB which is a consequence of the influence of the subregions of the probe face. By means of this technique we can predict the faces influenced by the horizontal subregions of the probe face by inheriting the PCA architecture described in the above section (3.5). Let the similarity measure between two faces F_i, F_j be denoted by $SM(F_i, F_j) \in [0,1]$. If $i = j$ then $SM(F_i, F_j)$ will be 1. Otherwise $0 \leq SM(F_i, F_j) < 1$. Faces influenced (Ξ) by k subsamples can be computed by

$$\Xi = \arg\min_{F_j} SM(X_{ij}, F_j), \qquad i = 1, \dots, k, \qquad j = 1, \dots, n. \tag{3.9}$$

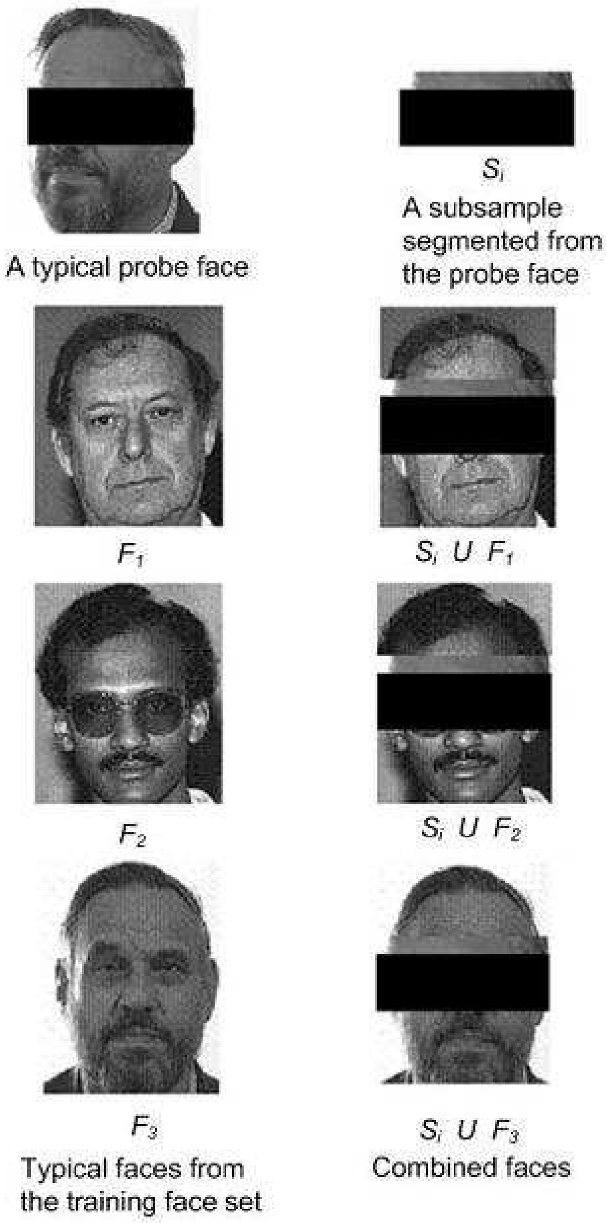


FIGURE 3.6: The process of generating combined faces from a subsample of a typical occluded face being illustrated; Face images from FERET [2] dataset has been used.

We construct the eigenspace (face space) off-line from the gallery set of face samples of the FDB. Since S_i are components of the probe face I , we have that

$$I = S_1 \cup S_2 \ldots \cup S_k \tag{3.10}$$

which mathematically conveys that a face image is a combination of its facial components (subsamples). The following equation is used to project the combined face X_{ij} into the eigen-space.

$$\omega_{ij} = u_j^T(X_{ij} - \Psi), \qquad i = 1, \dots, k, \qquad j = 1, \dots, n, \tag{3.11}$$

where ω_{ij} , u_j and Ψ are respectively the weight vectors, eigenvectors and the mean face of the FDB. The face space projection Φ_f can be computed by

$$\Phi_f = \sum_{j=1}^{n} \omega_{ij} u_j, \qquad i = 1, \dots, k. \tag{3.12}$$

The Euclidean distance between X_{ij} and the face space projection can be computed using

$$\epsilon_{ij} = \parallel (X_{ij} - \Psi) - \Phi_f \parallel \tag{3.13}$$

Let E_s represent the sorted Euclidian distances of ϵ_{ij}. Consequently the r face classes that correspond to the first r Euclidean distances of E_s will yield the faces influenced by each of the horizontal subregions of the probe face. Similar to how a human might recall some faces by observing portions of a face, the above formulation aids the machine to shortlist faces by observing subsamples of a face via psychophysical means.

3.6 Learning the Parameters of the BN from FDB

Since the FDB is readily available, the prior belief states of the subsamples, which are the parameters of the proposed BN, can be computed off-line before the probe face is observed. The belief states of a subsample S_i intuitively represent how effectively it can contribute to the recognition of faces. Dirichlet density functions are widely used in Bayesian statistics as they provide intuitive means in representing prior beliefs which can be updated gradually by observing evidence [147].

The prior belief states of the proposed BN can be quantified using the following Dirichlet density function

$$\rho(f_1, f_2, f_3, \ldots, f_{r-1}) = \frac{\Gamma(N)}{\Pi_{k=1}^{r}\Gamma(a_k)} f_1^{a_1-1} f_2^{a_2-1} \cdots f_r^{a_r-1} \tag{3.14}$$

where $f_1, f_2, f_3, \ldots, f_{r-1}$ are values of random variables $F_1, F_2, F_3, \ldots, F_{r-1}$, $0 \leq f_k \leq 1$, $\sum_{k=1}^{r} f_k = 1$, $a_1, a_2, a_3, \ldots, a_r$ are integers ≥ 1 and $N = \sum_{k=1}^{r} a_k$. The gama function used in (3.14) is computed by

$$\Gamma(x) = (x-1)! \qquad x > 0 \tag{3.15}$$

The prior belief states of the parameters which are the fundamental building blocks of the BN are updated by a machine learning procedure called parameter estimation. Out of several such procedures available, two of them, Maximum-Likelihood Estimation (MLE) and Bayesian estimation are considered most often by researchers [154]. When compared to Bayesian estimation, MLE is simpler. MLE has been recommended by [155] as it has many optimal properties in estimation including asymptotic consistency and unbiased nature. MLE demands large training samples. Fortunately as the BN can be realized through large samples available in the facial domains, MLE will converge to precise estimates enabling the distribution of the parameters to be normal. Consequently many of the inference methods in statistics such as Chi-square test, Bayesian methods, Akaike information criterion [156] and Bayesian information criteria [157] are developed based on MLE. Equation (2.6) reveals that the Markov condition has been satisfied by the probability distribution entailed by the DAG of the proposed PROCIM model. Hence we have

$$P(F|S_i) = \Pi_{j=1}^{n} P(F_j|S_i), \qquad i = 1, \ldots, k. \tag{3.16}$$

Recall from Section 3.4 that F represents the n faces in the gallery (training) set and S represents the k subsamples of the probe face. We mathematically define the influence strength Z_{ij} of a subsample S_i as

$$Z_{ij} = (n - \ell)/n \tag{3.17}$$

where ℓ is the rank in which the face F_j is being recognized by the subsample S_i. This clearly shows that F depends on Z_{ij}; without a gallery of faces, it would be impossible to define Z_{ij}.

The objective of MLE is to estimate the unknown parameter Z_{ij} that best agrees with the observed gallery set of face images. MLE of Z_{ij} is by definition the value of $\hat{Z}_{ij}$ that maximizes $lnP(F|S_i)$, the log likelihood of the parameter set Z_{ij} with respect to the training face set F. The log likelihood is dependent on Z_{ij}, but we do not show this, to simplify the notation. The parameter $\hat{Z}_{ij}$ can be computed by

$$\hat{Z}_{ij} = \arg\max_{Z_{ij}} \ \ln\, P(F|S_i). \tag{3.18}$$

To be a maximum, the shape of the log-likelihood function should be convex in the neighborhood of $\hat{Z}_{ij}$ which can be checked by computing the second derivatives of the log likelihoods. Note that

$$\ln P(F|S_i) = \sum_{j=1}^{n} \ln P(F_j|S_i). \tag{3.19}$$

With this, equation (3.18) becomes

$$\hat{Z}_{ij} = \arg\max_{Z_{ij}} \sum_{j=1}^{n} \ln P(F_j|S_i). \tag{3.20}$$

For large values of n, Z_{ij} defined in (3.17) becomes nearly continuous. The maximization in (3.20) can be performed by a gradient descent, provided that the numerical algorithm has a steplength that is not lower than the largest difference between two Z_{ij}, or that the algorithm uses a suitable interpolation routine. By using a gradient method, a set of necessary conditions for the maximum-likelihood estimate for Z_{ij} can be obtained from the set of k equations

$$\sum_{j=1}^{n} \nabla_{Z_i} \ln P(F_j|S_i) = 0 \qquad\qquad i = 1, \ldots, k \tag{3.21}$$

where the gradient operator ∇_{Z_i} is given by

$$\nabla_{Z_i} \equiv \begin{pmatrix} \frac{\partial}{\partial Z_{i1}} \\ \frac{\partial}{\partial Z_{i2}} \\ \vdots \\ \frac{\partial}{\partial Z_{in}} \end{pmatrix} \qquad\qquad i = 1, \ldots, k \tag{3.22}$$

Remember that Z_i is a vector with n components $Z_{ij}, j = 1, \ldots, n$. The equations above are similar to the standard equations for MLE, as can be found for example in [154]. The only difference is that the standard equations have a simple vector of parameters, and that our parameter vector Z_i is dependent on the subsample. Indeed, we perform a standard MLE, but we do it for all subsamples $i = 1, \ldots, k$. The BN employed in PROCIM learns the belief states of conditional probability potentials systematically from the training face images using the MLE approach outlined above. The learnt belief states are stored in the form of Conditional Probability Tables (CPTs). For k number of subsamples, the BN yields a CPT comprising $2^k - 1$ rows. A typical CPT for the case of $k = 5$ is shown in Table 3.1. The conditional probabilities in Table 3.1 give the likelihood measures of faces given their various subsample or subsample combinations being observed. Usually when all the subsamples together are observed, the probability of faces being recognized is larger. But this general behavior is not always applicable and varies from individual to individual, as different individuals can be characterized by a specific combination of facial features. For example, by referring to the last column of the table, we observe that conditional probabilities $P(F10|S1)$ and $P(F10|S1, S3)$ are relatively higher. This reveals the fact that the face recognition of subject $F10$ is highly characterized by a specific subregion or a specific combination of subregions. However the conditional probabilities $P(F10|S4)$, $P(F10|S5)$ and $P(F10|S4, S5)$ being low indicate that subject $F10$ is poorly characterized by these subregion(s).

Table 3.1: Learnt belief states that represent the likelihood of typical faces (chosen from FERET dataset) by observing subsample combinations for $k = 5$

	F1	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**
S1	0.933	0.950	0.950	0.633	0.392	0.950	0.917	0.158	0.792	0.933
S2	0.933	0.883	0.950	0.317	0.233	0.900	0.792	0.775	0.633	0.600
S1S2	0.950	0.950	0.950	0.600	0.317	0.950	0.883	0.742	0.758	0.792
S3	0.342	0.500	0.850	0.517	0.425	0.900	0.633	0.583	0.708	0.375
S1S3	0.950	0.758	0.950	0.475	0.583	0.950	0.917	0.883	0.633	0.917
S2S3	0.583	0.708	0.950	0.533	0.600	0.950	0.792	0.867	0.633	0.583
S1S2S3	0.950	0.867	0.950	0.567	0.675	0.950	0.900	0.850	0.633	0.775
S4	0.125	0.917	0.708	0.375	0.125	0.867	0.633	0.758	0.775	0.125
S1S4	0.692	0.950	0.950	0.617	0.408	0.950	0.883	0.933	0.633	0.775
S2S4	0.408	0.917	0.917	0.533	0.408	0.950	0.792	0.792	0.633	0.500
S1S2S4	0.775	0.950	0.950	0.550	0.500	0.950	0.792	0.883	0.742	0.775
S3S4	0.125	0.792	0.850	0.300	0.267	0.950	0.742	0.850	0.742	0.233
S1S3S4	0.817	0.950	0.950	0.617	0.517	0.950	0.792	0.933	0.633	0.775
S2S3S4	0.300	0.917	0.950	0.500	0.533	0.950	0.792	0.867	0.633	0.408
S1S2S3S4	0.850	0.933	0.950	0.550	0.583	0.950	0.792	0.883	0.633	0.775
S5	0.567	0.900	0.725	0.108	0.125	0.775	0.775	0.317	0.883	0.125
S1S5	0.950	0.950	0.933	0.550	0.683	0.950	0.933	0.425	0.775	0.883
S2S5	0.933	0.933	0.900	0.533	0.283	0.933	0.933	0.458	0.758	0.517
S1S2S5	0.950	0.950	0.933	0.625	0.300	0.950	0.950	0.425	0.758	0.867
S3S5	0.625	0.792	0.883	0.567	0.342	0.933	0.758	0.408	0.758	0.408
S1S3S5	0.950	0.950	0.933	0.442	0.442	0.950	0.933	0.692	0.742	0.867
S2S3S5	0.850	0.917	0.917	0.583	0.500	0.950	0.775	0.642	0.633	0.533
S1S2S3S5	0.950	0.950	0.933	0.583	0.767	0.950	0.950	0.550	0.725	0.850
S4S5	0.267	0.917	0.725	0.533	0.125	0.917	0.883	0.408	0.883	0.125
S1S4S5	0.883	0.950	0.933	0.533	0.425	0.950	0.933	0.658	0.742	0.833
S2S4S5	0.642	0.950	0.917	0.583	0.283	0.950	0.917	0.567	0.725	0.392
S1S2S4S5	0.917	0.950	0.933	0.300	0.425	0.950	0.950	0.658	0.758	0.758
S3S4S5	0.300	0.900	0.883	0.283	0.142	0.950	0.792	0.317	0.758	0.342
S1S3S4S5	0.917	0.950	0.933	0.583	0.442	0.950	0.933	0.817	0.633	0.742
S2S3S4S5	0.500	0.950	0.917	0.517	0.517	0.950	0.883	0.533	0.633	0.408
S1S2S3S4S5	0.900	0.950	0.933	0.583	0.625	0.950	0.900	0.642	0.742	0.742

Thus the proposed framework allows to characterize subtle subsample relationships and exploit them to mitigate the vast uncertainties imposed by occlusions and other common variations.

3.7 Deciding on the most Probable Faces

In section 3.5.1, we have formulated a procedure to shortlist a reduced set of faces influenced by horizontal subsamples of a face image from the huge FDB. Here we will propose a formula to rank-list these shortlisted faces in order to decide the most probable (winner) faces. Say C number of faces have been influenced for a given probe face after the horizontal subregions have been processed. The formula which we intend to formulate will yield a score for each of the C number of faces. This score will aid to rank-list the faces.

By exploiting the graphical structure of the BN, the probability distribution over $F_m, 0 < m \leq C$ can be computed by

$$P(F_m) = \sum_{S_i} P(S_i)P(F_m|S_i) \qquad i = 1, \ldots, k \tag{3.23}$$

where $P(S_i)$ is the prior probability of subsamples and $P(F_m|S_i)$ is the probability of a face given the condition that some subsamples (or a subsample) has influenced it.

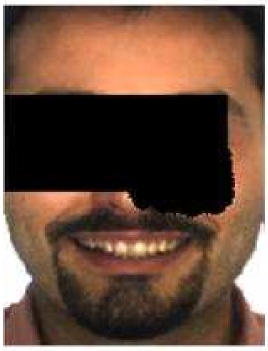


FIGURE 3.7: A typical occluded probe face from AR face set

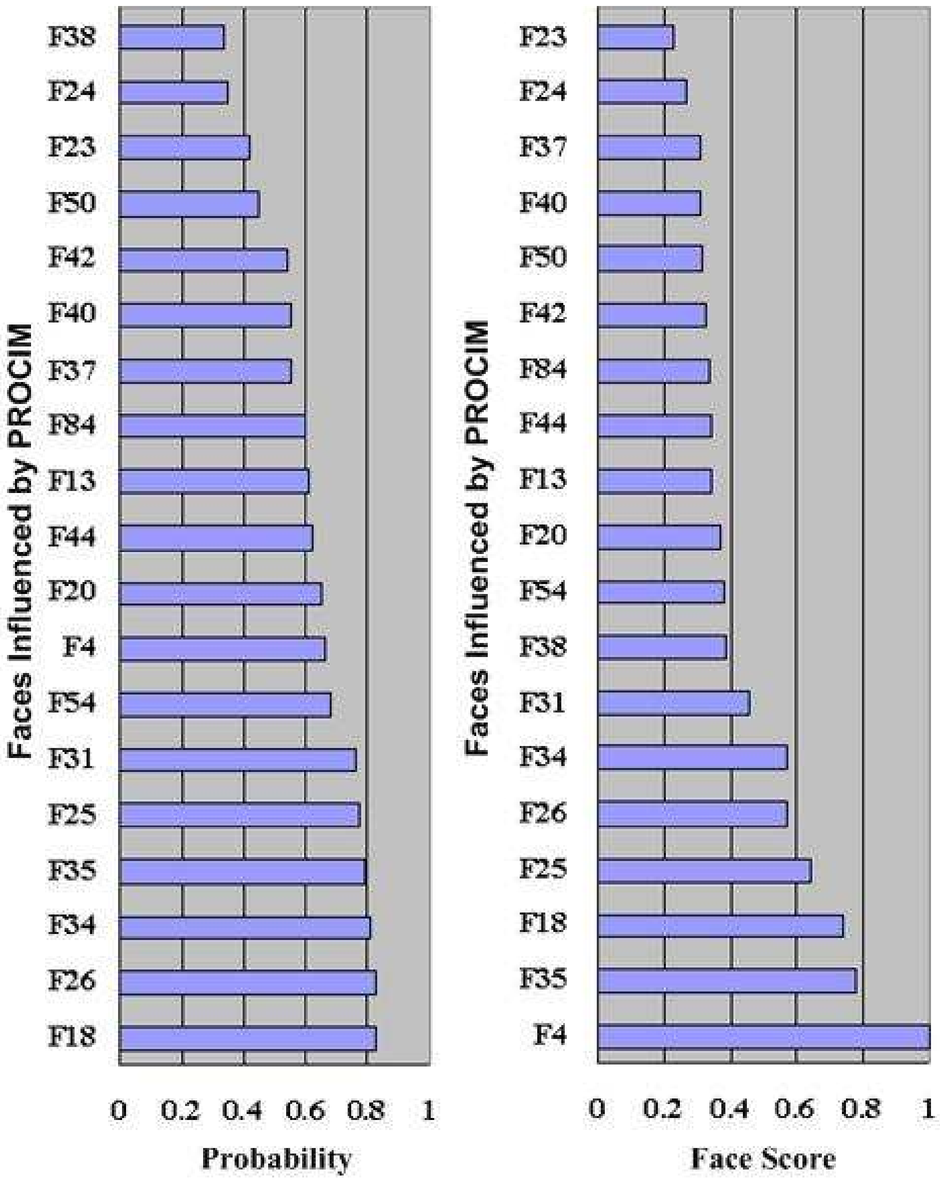


FIGURE 3.8: Comparison of Probability Distribution Vs. Face Score; Mere probability theory is not adequate to make a meaningful decision; The proposed Face Score formula which is based on probability as well as utility theory discriminates the winner face better than a formula which just uses probabilities.

Consider the case where an occluded probe face shown in Fig. 3.7 has been subjected to similarity mapping based processing. By applying the procedures formulated in section 3.5.1 faces have been short listed. The probability distribution of these faces are shown in the left most bar chart of Fig.3.8. We see that the probability of F4 (the gallery instance of the probe) falls somewhere in the middle. Also the probabilities of faces (F18,F26,...,F54) which are higher than the probability for F4, fetch similar values without much discrimination. This reveals that mere probabilities are not adequate enough to make a meaningful decision. To counter this problem we consider the well known rule of thumb given by Russel and Norvig [140] which emphasizes that "Probability theory and utility theory together constitute decision theory". By utilizing the crucial influence strengths Z defined in Section 3.4 to weigh the prior probability of subsamples, logically the face score will yield meaningful results if it is a function of the following two vital factors,

i. The probability distribution of the faces and their subsamples.

ii. Weighted influence strength between faces influenced and their subsamples.

Consolidating the above factors, the face score μ of a mth face can be computed using

$$\mu(F_m) = \sum_{S_i} P(S_i)P(F_m|S_i) + \sum_{S_i} Z_{im}P(S_i) \tag{3.24}$$

If a subsample has not influenced a face, due to severe uncertainty, Z would fetch a zero. This will inturn nullify the face score if we use a product operation. Hence the face score appropriately uses an addition operation. With the aid of this face score, faces have been rank-listed as shown by the right most bar chart of Fig.3.8. The chart shows that the face score discriminates the winner faces well and the probe face has been well classified (rank 1). The decision process involved in MLA [125] relies on the probability of a given local match which is directly associated with a distance metric. However, the decision process employed in the proposed PROCIM model makes use of the influence strength in conjunction with the probabilities entailed in the model. Such a consolidated decision process enables the model to gain discrimination power.

3.8 Understanding Similarity Mappings by Visualizing the Model

Though humans can communicate sensibly by establishing and exploiting similarity mappings in a knowledge domain, it may be difficult to fully understand its philosophical aspects. Visualizing the graphical structure of the proposed model enables the reader to understand the abstract similarity mapping phenomenon intuitively. The three instances of a typical probe face being subjected to Nil, Minor and Major occlusions are shown in Fig.3.9.

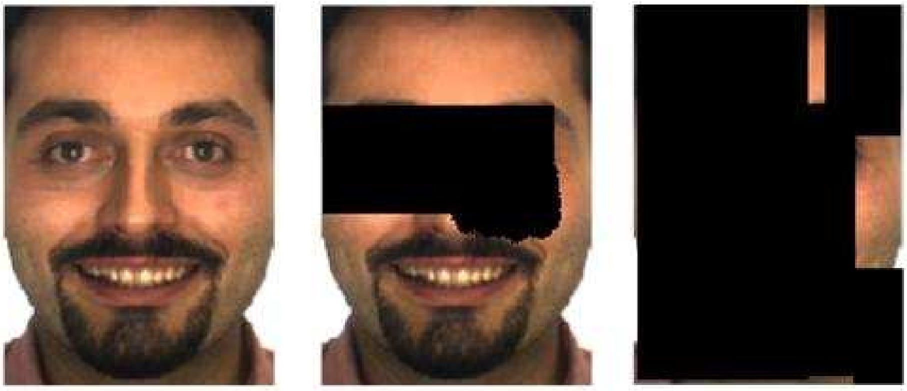


FIGURE 3.9: A typical probe face being subjected to nil, minor and major occlusion has been taken to study the relationships between occlusion and similarity maps

The corresponding BNs generated by PROCIM and the bar chart showing Face Score Vs. Rank-listed Faces are shown in Figures 3.10,3.11,3.12,3.13,3.14 and 3.15. When there are no occlusions, the gallery instance F4 of the given probe is mapped by all the subsamples as shown in Fig.3.10. As the occlusion content increases the mappings tend to reduce (less number of subsamples map to F4 as shown in Fig.3.12 and Fig.3.14. This shows that there exists strong relationships between occlusions and similarity mappings. Occlusions are capable of muddling intrinsic similarity relationships that exists in facial domains.

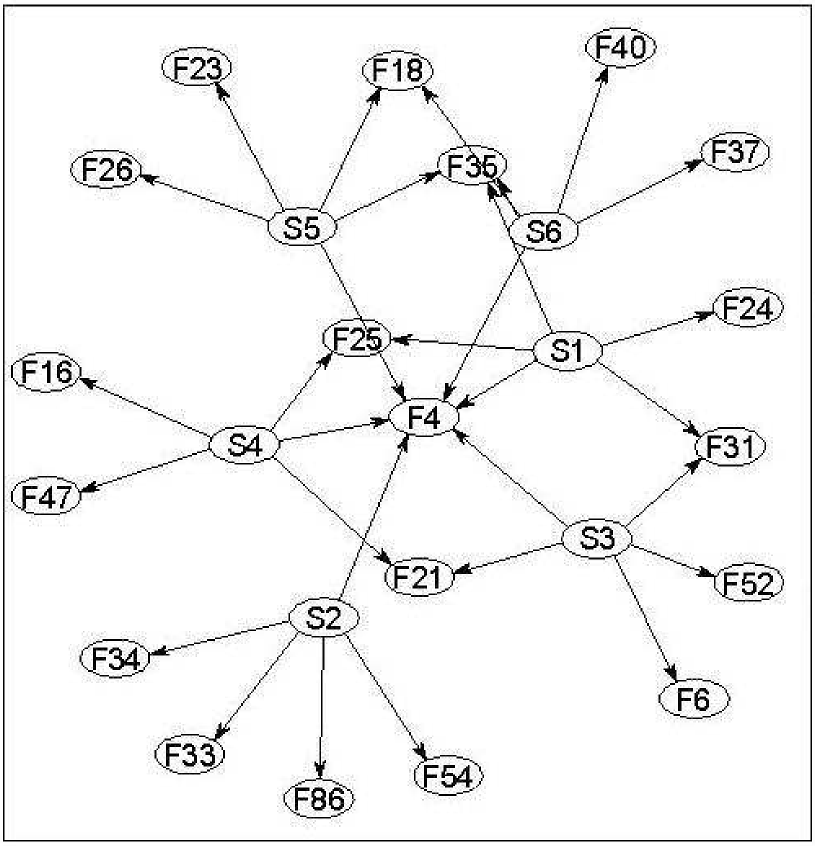


Figure 3.10: BN generated by PROCIM for the case of probe face with nil occlusion; When the probe face is not prone to occlusion, its gallery instance $F4$ is mapped by more subsamples

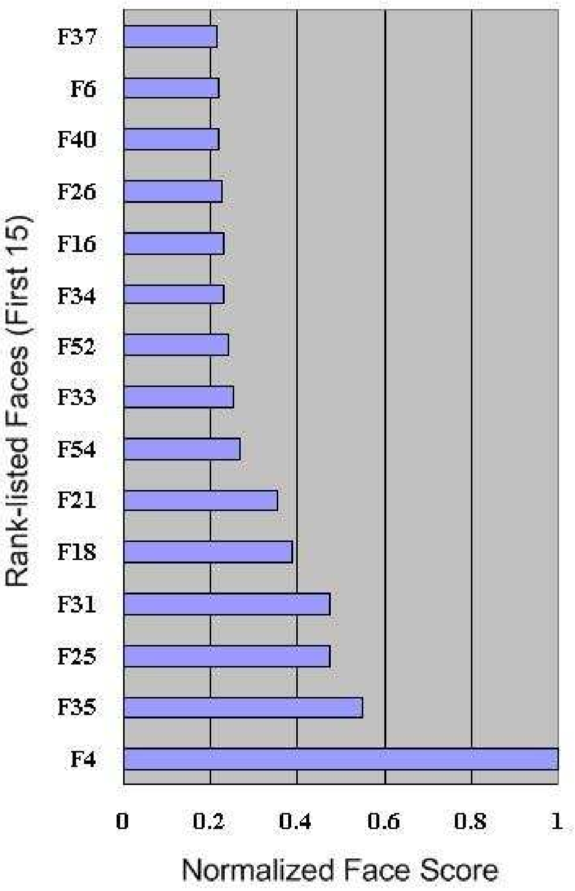


FIGURE 3.11: Rank-listed faces for the case of recognizing the probe face with nil occlusion; The winner subject $F4$ is well discriminated from other faces

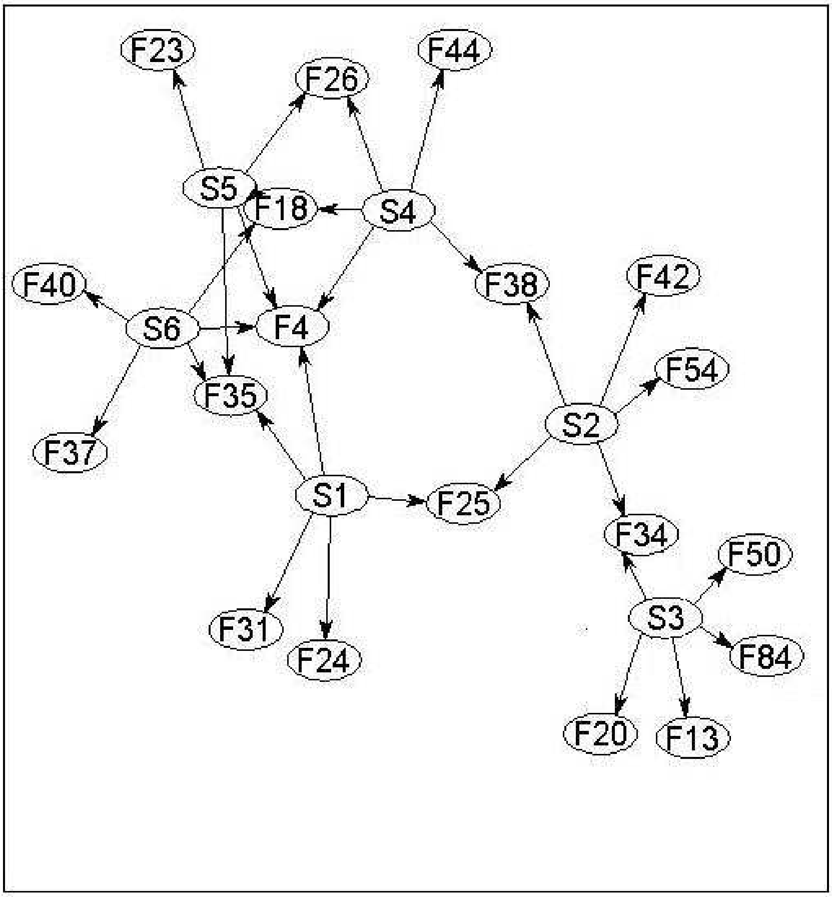


FIGURE 3.12: BN generated by PROCIM for the case of probe face with minor occlusion; When compared to the BN shown in Fig. 3.10, $F4$ is mapped by less number of subsamples

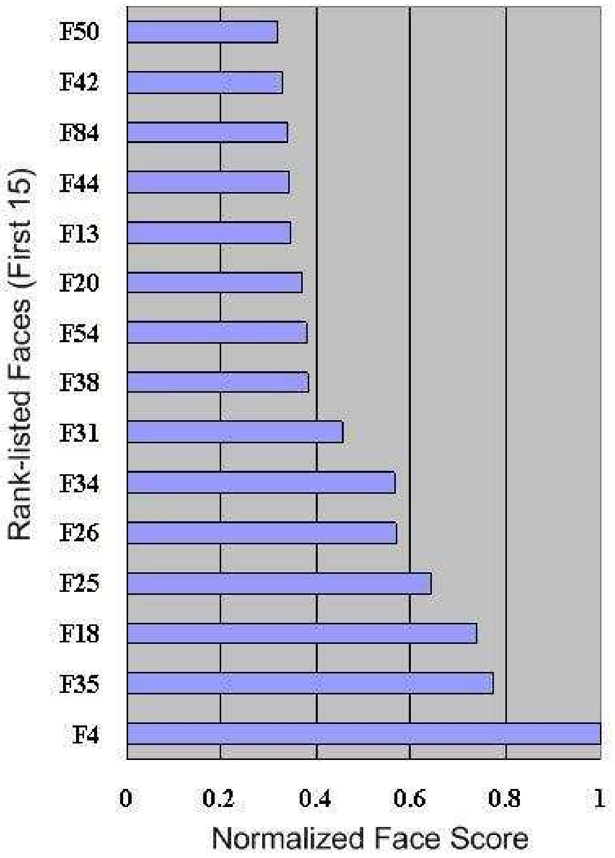


FIGURE 3.13: Rank-listed faces for the case of recognizing the probe face with minor Occlusion; Despite the presence of minor occlusion, still $F4$ is well discriminated by the proposed Face Score formula

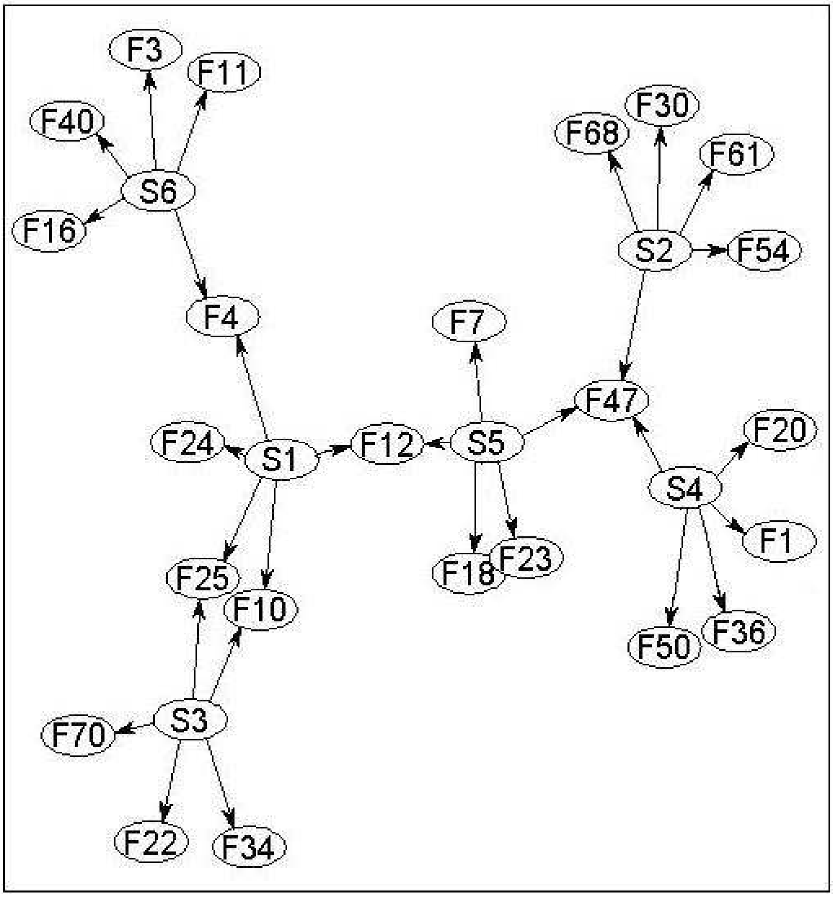


Figure 3.14: BN generated by PROCIM for the case of probe face with major Occlusion; When compared to the BNs shown in Fig.3.10 and Fig.3.12, $F4$ is mapped by very less number of subsamples; This shows that occlusions are capable of muddling similarity mappings

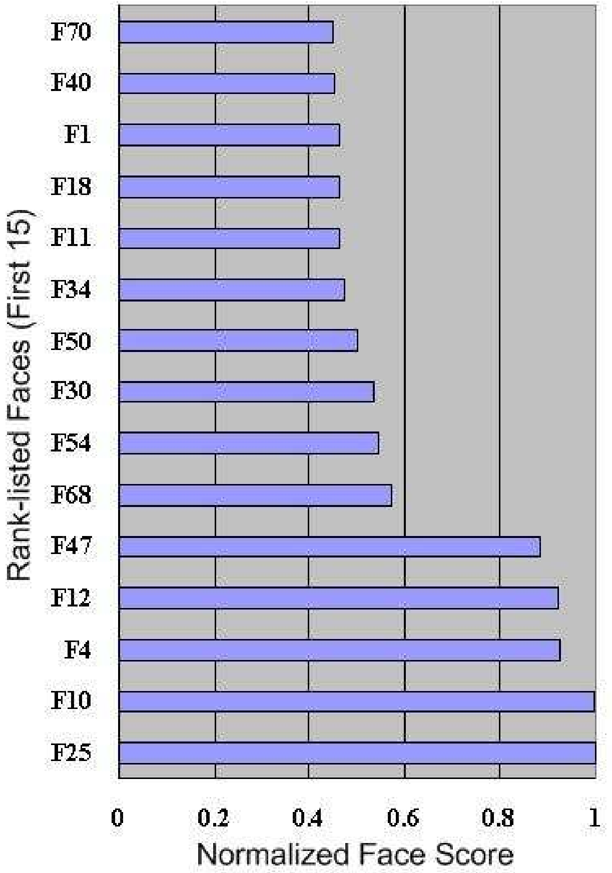


FIGURE 3.15: Rank-listed faces for the case of Recognizing the Probe Face with Major Occlusion; As a consequence of major occlusion $F4$ has been pushed from rank 1 to rank 3;

PROCIM learns vital information about the likelihood measures of various subsamples and the faces that encode them and store them systematically in conditional probability tables. Further during the training process PROCIM learns the prior belief states of subsamples. For the case of faces rank-listed while recognizing the face with major occlusions shown in Fig. 3.9 the conditional probabilities that have been learnt from the training data set are shown in Fig. 3.16. Due to the uncertainty caused by major occlusions only few subsamples contribute to the recognition of $F4$ as shown by the BN in Fig. 3.14. Further the influence strength achieved by $F47, F10, F25$ and $F12$ are higher than $F4$ as seen in Fig.3.17. But the conditional probabilities (likelihoods) and the prior belief states (Please see Fig.3.16 and Fig.3.18) of subsamples which have been learnt from the training dataset compromise the strengths diluted by occlusions. Consequently $F4$ is reasonably classified as seen in Fig.3.15. This intelligent trade off between the learnt information and observed mappings enables PROCIM to recognize faces with major occlusions.

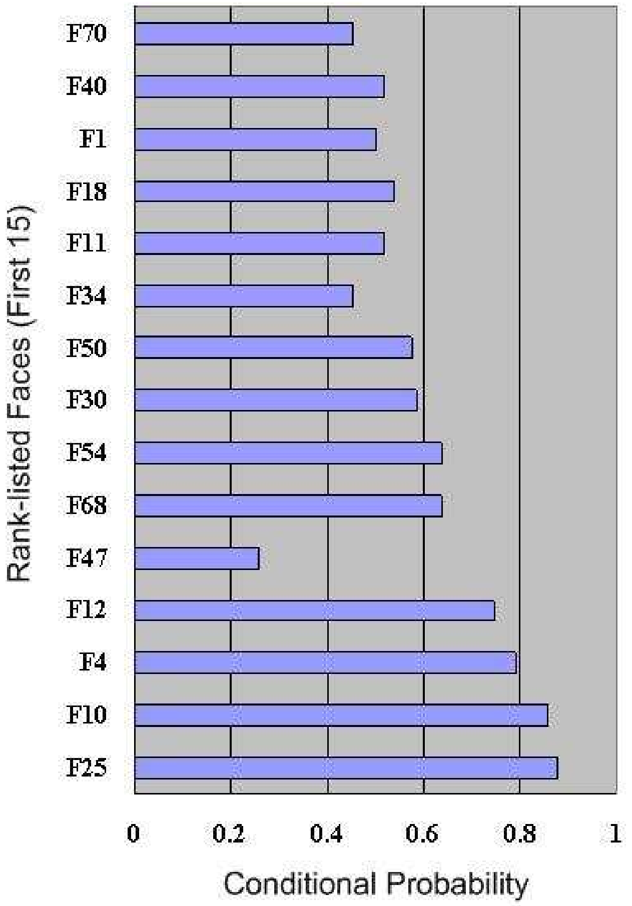


FIGURE 3.16: Represent the conditional probability potentials learnt from the AR training dataset for the first 15 winner faces associated with Fig. 3.14

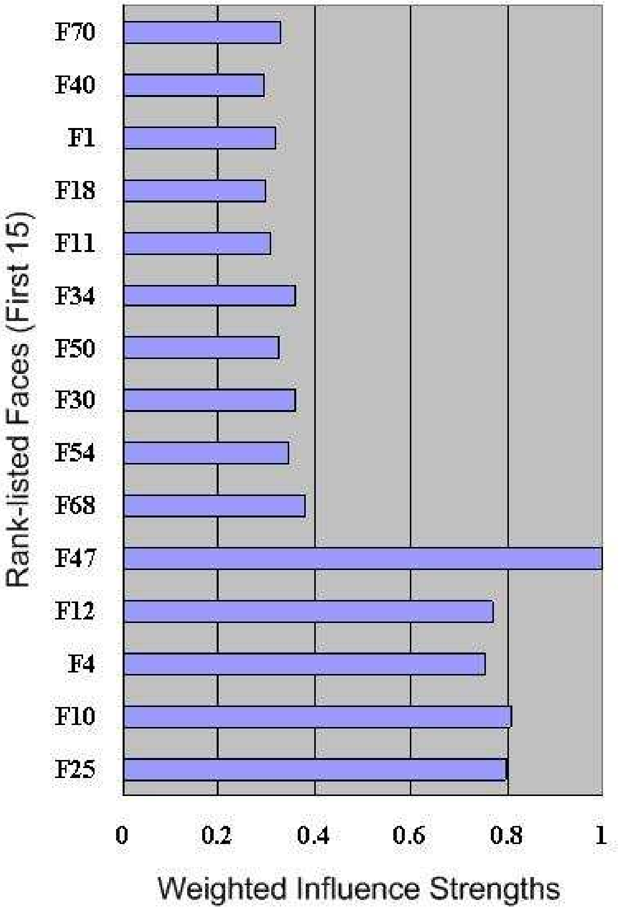


FIGURE 3.17: Corresponding influence strengths observed from the graph which have been duly weighted by priors for the first 15 winner faces associated with Fig. 3.14

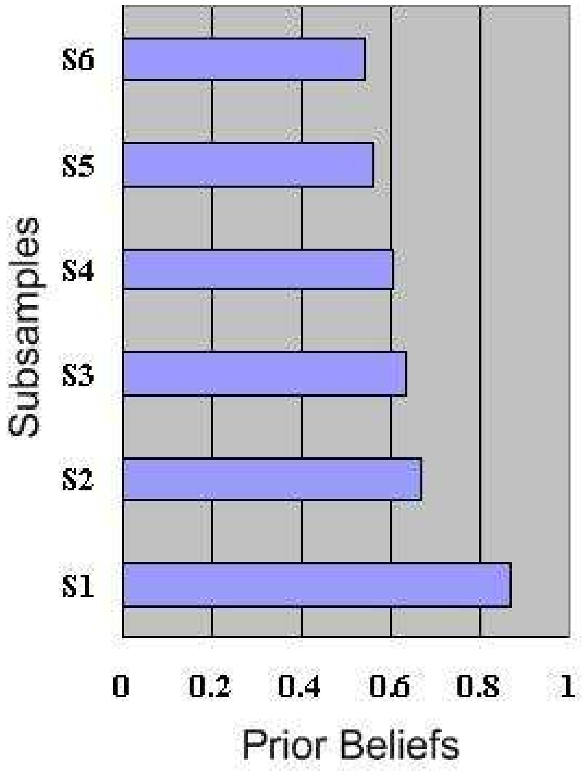


FIGURE 3.18: Prior belief states of subsamples learnt from data; As a face contains diverse source of information at various subregions, naturally the belief states of these subregions are also different

To mitigate the localization errors, MLA [125] attempts to find the subspace within the eigen-space where the localization error is minimal. The abstract nature of learning of such a subspace for each of the components and how well all these component based processing is collectively represented, is hard to visualize and interpret in the MLA approach. PROCIM learns and exploits intrinsic similarity relationships that are inherent in the facial domain to tackle the uncertainties. The graphical nature of the PROCIM enables us to visualize the similarity mappings inherent in various subsamples and how they jointly contribute to the recognition mechanism. Furthermore, how they gradually vary with respect to the presence of varied degrees of occlusions is clearly demonstrated using the proposed BN oriented

PROCIM architecture. This transparency is due to the intuitive psychophysical nature of the model.

3.9 Evaluation of PROCIM

We have implemented the proposed PROCIM model using MATLAB on Intel Pentium IV Core 2 Duo 2.39Ghz CPU with 2GB of RAM. We have made use of the routines offered by [158] and [159] to build the PCA feature space and BN respectively. We have evaluated the performance of PROCIM model using a series of experiments on standard datasets, the details of which are given in Section 3.9.1. As the nature of the underlying problem is identification and not verification, the performance of PROCIM is evaluated using a closed universe model which insists that all the probes need to have a match in the gallery. Such an evaluation model allows us to ask "How good is an algorithm at identifying a probe image?" [2]. The emphasis here is not always "is the top match correct?" but "is the correct answer in the top n matches?". To consider this vital emphasis, we report the performance statistics as recognition rates and Cumulative Match Characteristics (CMCs). CMC is a measure of identification performance which shows rank order statistics. In other words CMC indicates the probability that the gallery subject will be among the top n matches, for a given probe.

3.9.1 Details of Face Databases (FDBs) used in our Experiments

As a general practice in pattern recognition, it is accepted that using at least 10 times as many training samples per class as the number of features is a good practice to follow [160]. This ratio needs to be larger for more complex classifiers [161]. Based on these guidelines we have used the following widely used FDBs in order to evaluate and compare the PROCIM model with other standard techniques. Fig.3.19 shows some sample images of these FDBs.

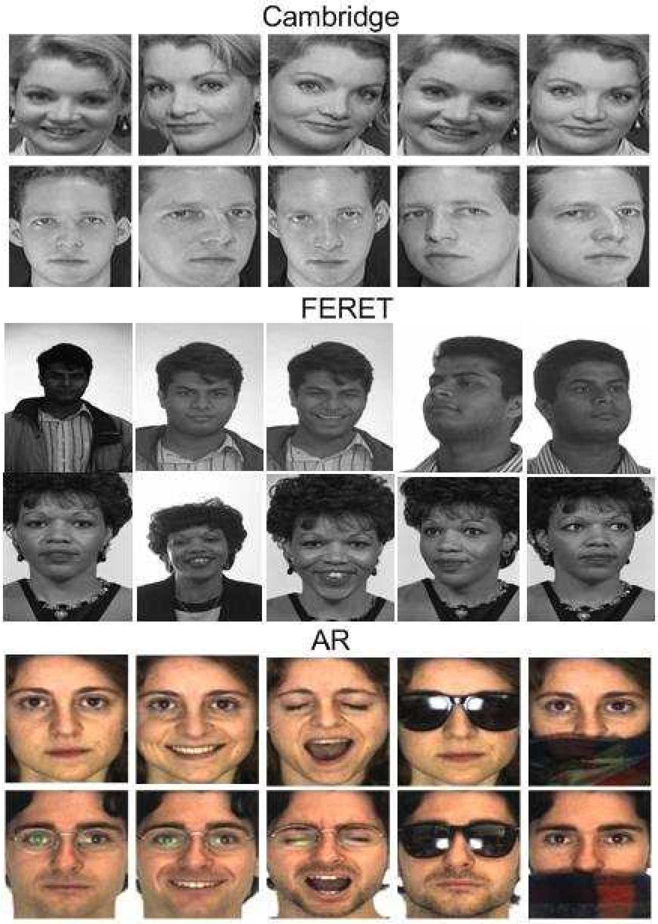


Figure 3.19: Sample face images typically chosen from the AT&T, AR and FERET FDBs

3.9.1.1 AT&T FDB

Initially we have used this FDB which is also known as AT&T FDB [62]. This FDB was formerly known as "The ORL Database of Faces", provided by AT&T Laboratories of Cambridge. We have used this FDB in order to compare our results with [114]. There are ten different images varying in scale, pose and expression for each of the 40 distinct subjects available in the FDB. The images are taken against a dark homogeneous background. In our experiments 50% of the face images were reserved as gallery set and the rest 50% were used as probe set. In other words gallery set is the set of known faces used for training, whereas the faces used to test the model are known as probe set.

3.9.1.2 AR FDB

We have used the huge AR FDB [124] which consists of over 3200 color images of 126 subjects. Images feature frontal view faces with different facial expressions, illumination conditions, and realistic occlusions (sun glasses and scarf). Each person participated in two sessions, separated by two weeks time. Per subject we have used four frontal view images chosen at random for training.

3.9.1.3 FERET FDB

To investigate the generalization ability of the proposed model we have used the FERET (FacE REcognition Technology) FDB [2, 162] which contains face images collected under the FERET program sponsored by the DOD counter drug Technology Development Program Office. It is managed by the Defense Advanced Research Projects Agency (DARPA) and the National Institute of Standards and Technology (NIST). The FERET FDB has enabled researchers to develop and evaluate algorithms on a common large database of facial images that was gathered independently from the algorithm developers. This FDB has been designed to advance the state of the art in face recognition and as such face images were taken in real world settings in order to simulate typical unconstrained scenarios. Face images are subject to diverse variations such as pose, illumination, scale and rotation. For our experiments we have taken a subset of FERET consisting of 2184 face images with 8 variations per subject. We use five face images for training and three for testing. The training and test images have been chosen at random for each subject. We categorize the images into three different scenarios as summarized in Table 3.2. The term *duplicate* in the table, in the context of biometrics, refers to the probe image of a person whose corresponding gallery image was taken from a different image set. Usually, a duplicate is taken on a different day than the corresponding gallery image.

Table 3.2: FERET FDB Details

Category	**FERET Notation**	**Description**
Dataset A	Fa & Fb	Frontal view images including duplicates with variations in expression, illumination and scale
Dataset B	Above + ql & qr	Above variations + pose variations (rotation)
Dataset C	Above + b-series	Above variations + different settings of camera and lighting

3.9.2 Performance Evaluation

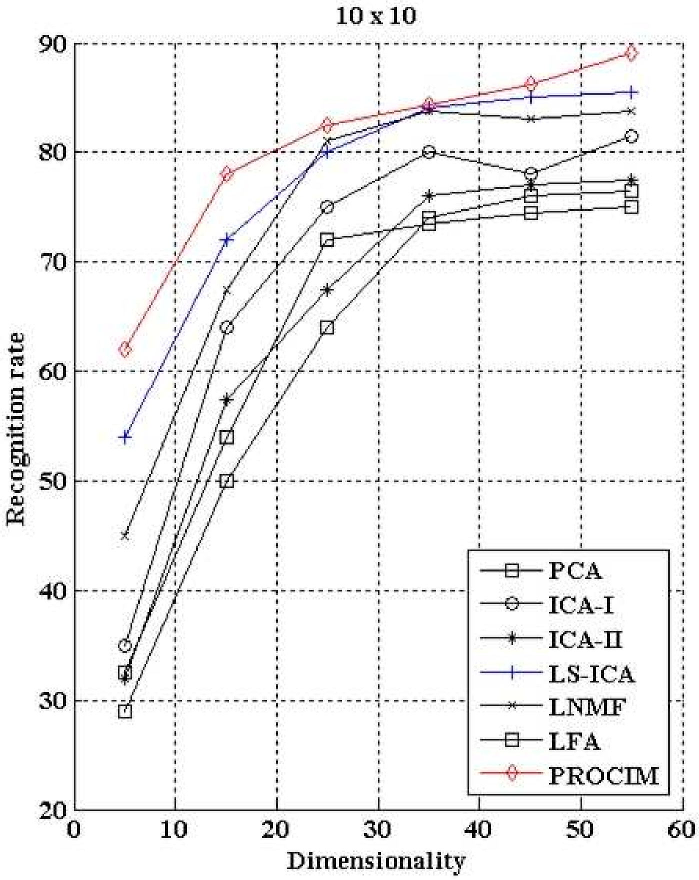


Figure 3.20: PROCIM compared with state-of-the-art techniques; Occlusion patch size: 10 × 10 pixels

Papers on recognizing occluded faces, such as [114, 123, 125, 132], suggest to simulate synthetic occlusions at random locations on the probe set of face images in terms of continuous white or dark (square/rectangular) patches. We have termed this test as Continuous Random Occlusion Test (CROT). Kim et al. [114] have taken into consideration occluded patch sizes from 10 x 10 to 30 x 30 to evaluate their proposed LS-ICA method with other techniques viz., PCA [26], ICA I & II [31], LNMF[135] and LFA[133]. We have compared the recognition performance of the proposed PROCIM model with these techniques as well using the AT&T FDB and projected the results in Figures 3.20, 3.21 & 3.22. The x-axis in these figures represent the dimensionality, that is the number of principal components that had been taken into consideration.

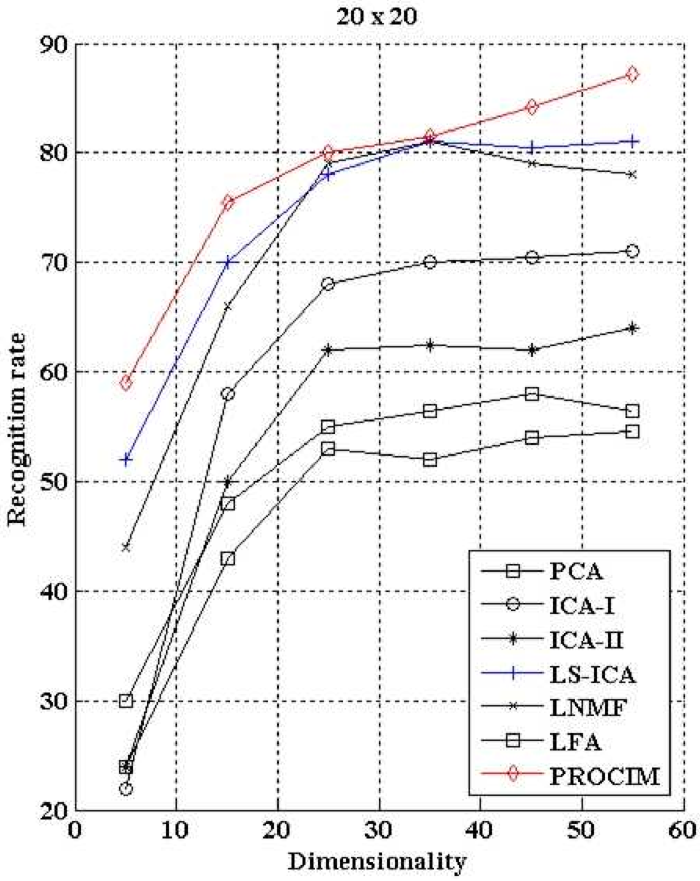


FIGURE 3.21: PROCIM compared with state-of-the-art techniques; Occlusion patch size: 20 × 20 pixels

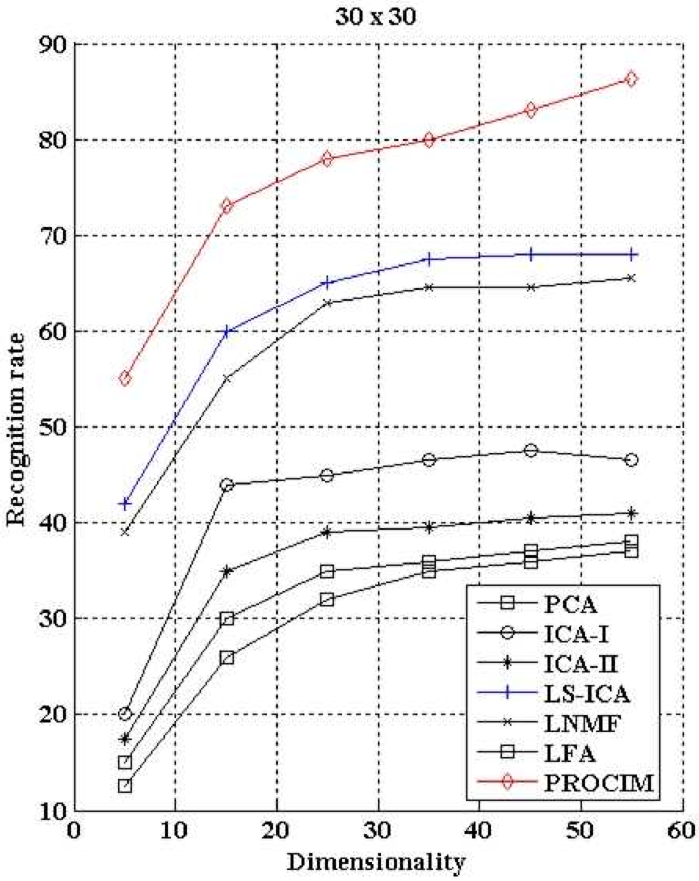


FIGURE 3.22: PROCIM compared with state-of-the-art techniques; Occlusion patch size: 30 × 30 pixels

It can be seen that PROCIM outperforms other techniques. While the other techniques degrade gradually as the occlusion content increases, PROCIM is stable. This stability is because of the fact that it can efficiently establish and exploit the similarity mappings inherent in facial domains to counter the uncertainties imposed by the occlusions.

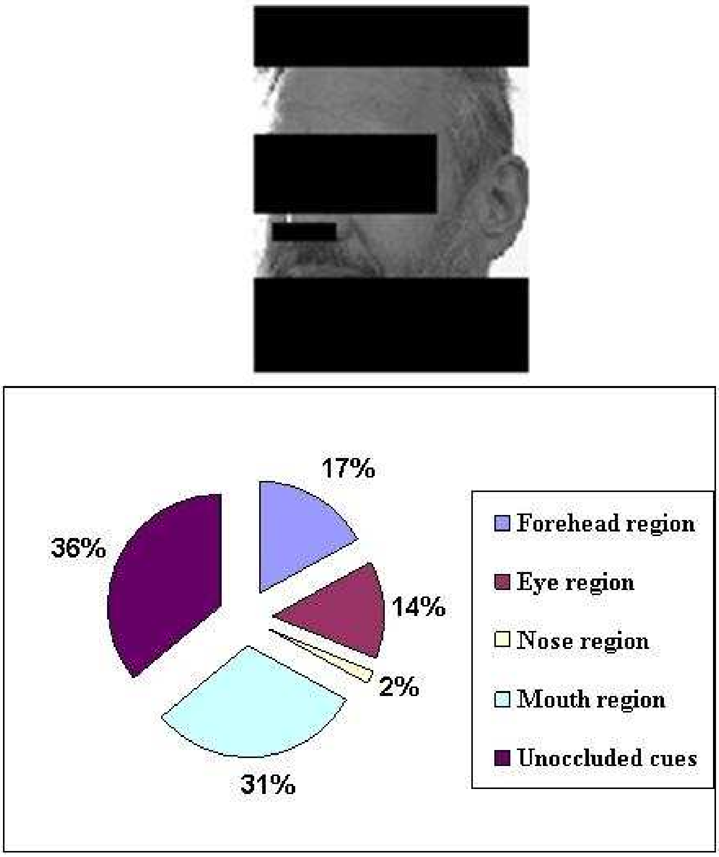


Figure 3.23: The Discrete Random Occlusions Test(DROT); In unconstrained scenarios, face images could be prone to multiple discrete occlusions and DROT simulates this reality

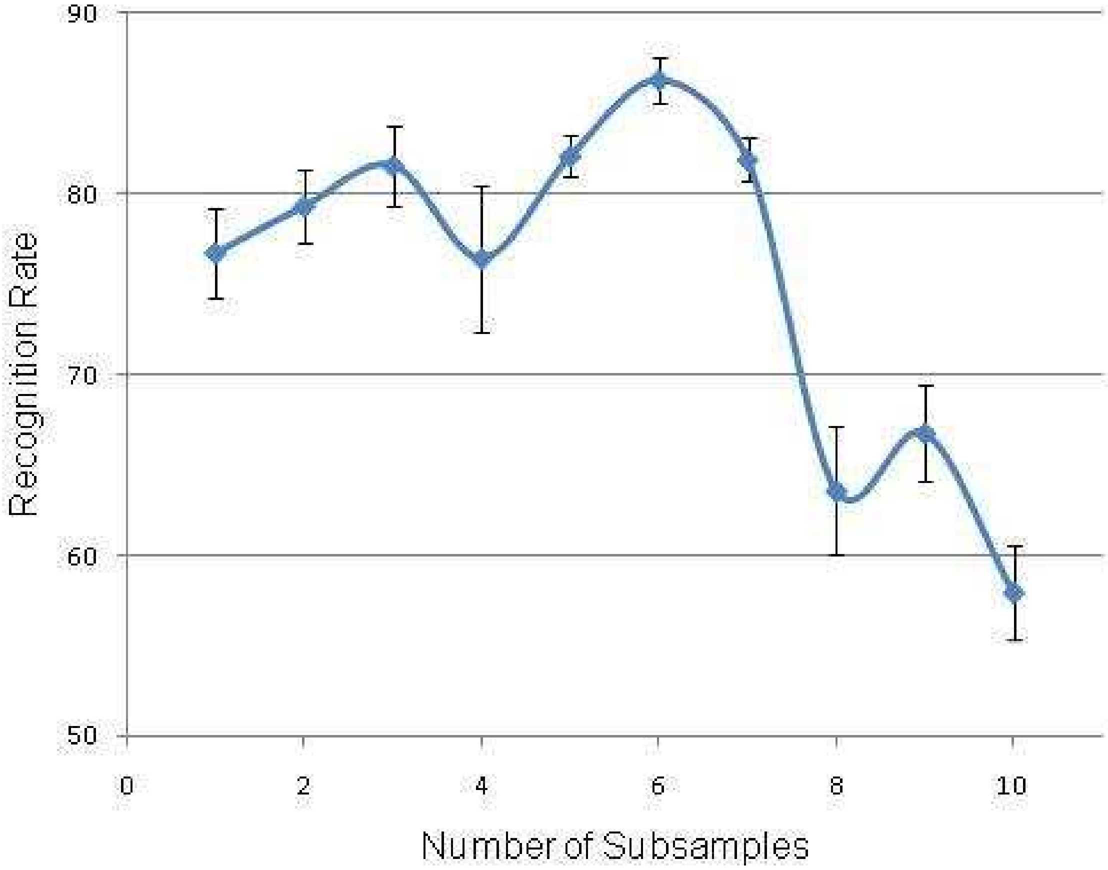
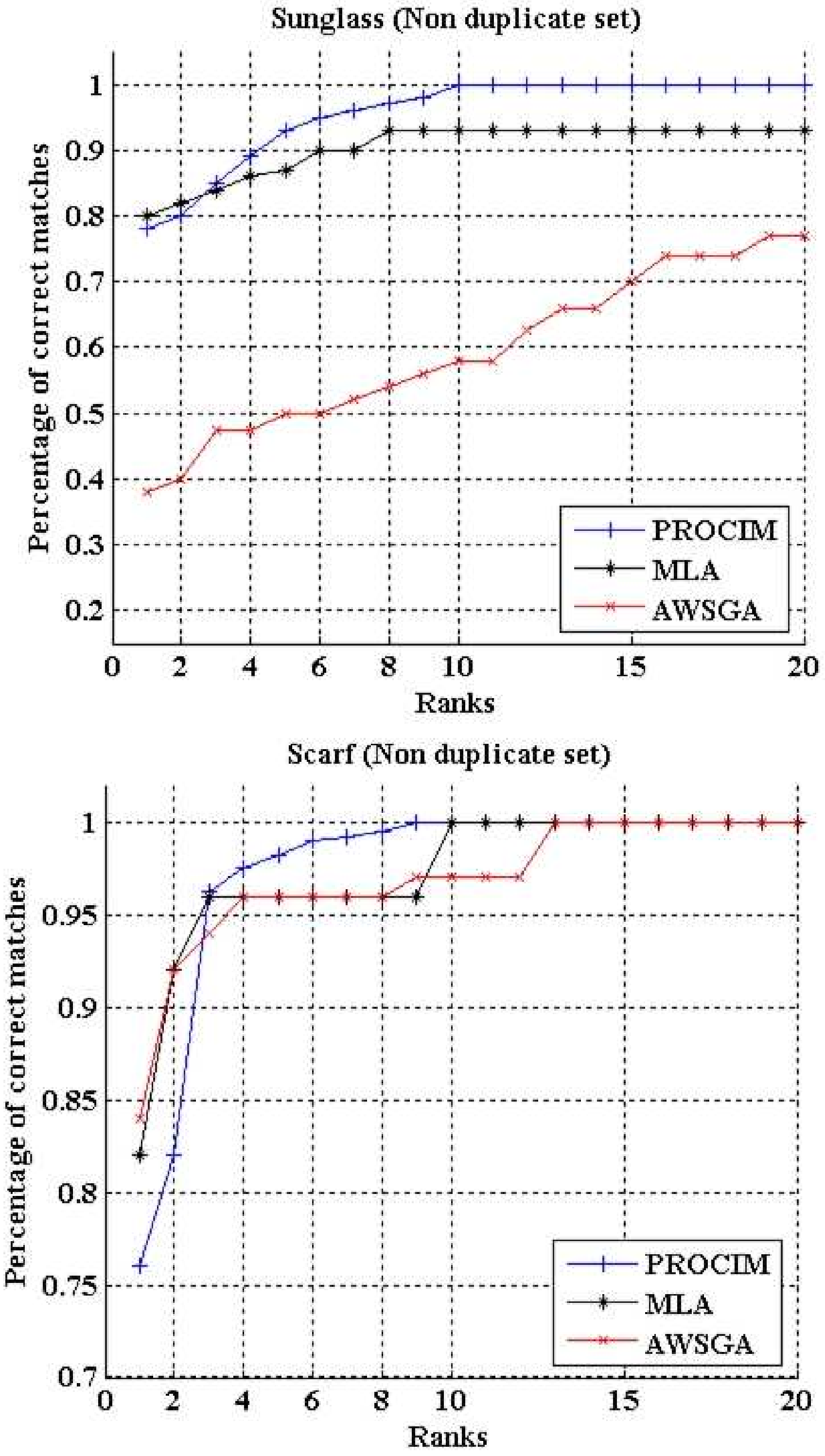


FIGURE 3.24: Estimating optimal number of subsamples using DROT; Few number of subsamples are sufficient to get optimal recognition performance. This means less computing resources are sufficient to run PROCIM. Half the length of the error-bars represent $\sigma/\sqrt{N}$, the standard-error, where σ is the standard-deviation of recognition rates for N number of trial-runs.

Contrary to CROT, in reality not always a single square occludes a face. For example a person could cover his face with a cap and sun glasses and not necessarily the entire upper face or lower face. To counter such multiple discrete occlusions we have simulated dark patches over various portions of the probe face set at random locations as shown in Fig.3.23 and performed recognition tests. We have named this test as "Discrete Random Occlusions Test (DROT)". Firstly we have used DROT to estimate the optimal number of subsamples required to achieve peak recognition performance. Secondly a face recognition model which can pass such an intuitive DROT will also be successful in recognizing faces with real occlusions such as cap, sun glasses and scarf possibly covering the face one at a time. The pie chart in Fig.3.23 quantifies these complex occlusions. Occlusion content of about 60% cover the probe face sets and only about 40% of cues are left out. The challenge is to exploit the minimal information (cues) available and predict winner faces. Initially DROT had been performed on the FERET dataset A by iterating the number of subsamples which will have an impact on the parameter size of the BN. From the graph shown in Fig.3.24 it is evident that by segmenting the face into just a few number of subsamples peak recognition rates can be achieved. Consequently this will minimize the usage of computing time and resources considerably. Though major occlusions are present, PROCIM has yielded a promising recognition rate of 86.3% with minimal number of parameters against DROT.

Further, the error bars in the graph indicate the measure of uncertainty inherent in the estimated subsamples with respect to the recognition rates. The error bars shown in the graph represent a description of confidence that the mean represents the true recognition rate with respect to the number of subsamples. The estimates where the error bars are shorter is an indication that the confidence levels at these estimates are higher and vice versa. We see that when the number of subsamples are between five to seven, the confidence level tend to be higher than the rest of the subsamples.

We have compared PROCIM with the very recently proposed Adaptively Weighted Sub-Gabor Array(AWSGA) approach [126] and Martinez's Localization Algorithm (MLA) [125] on the AR FDB where face images have been subjected to two types of real occlusions namely sunglass and scarf. Except the first two ranks, in the majority of the tests, PROCIM eventually outperforms both MLA and AWSGA as shown by the CMC graphs in Fig.3.25 and Fig.3.26. For example, referring to the graphs of Sunglass experiments, PROCIM reaches more than 90% recognition rate within ranks 5 and 7 respectively for the non-duplicate and duplicate test sets. But for MLA this happens only at ranks 8 and 11 respectively and for AWSGA it never happens even before 20 ranks, though both of them have the advantage of using one training sample per class. This means that PROCIM can recognize the actual suspect within less range of ranks which is crucial in criminal investigations, though more training samples per class are required. It is reported in [2, 125] that recognition tests on non duplicate images are tougher. Even against these tougher tests, PROCIM reports promising recognition rates of about 90% within 6 ranks.


Sunglass (Non duplicate set)

Scarf (Non duplicate set)

FIGURE 3.25: PROCIM versus MLA non-duplicate set

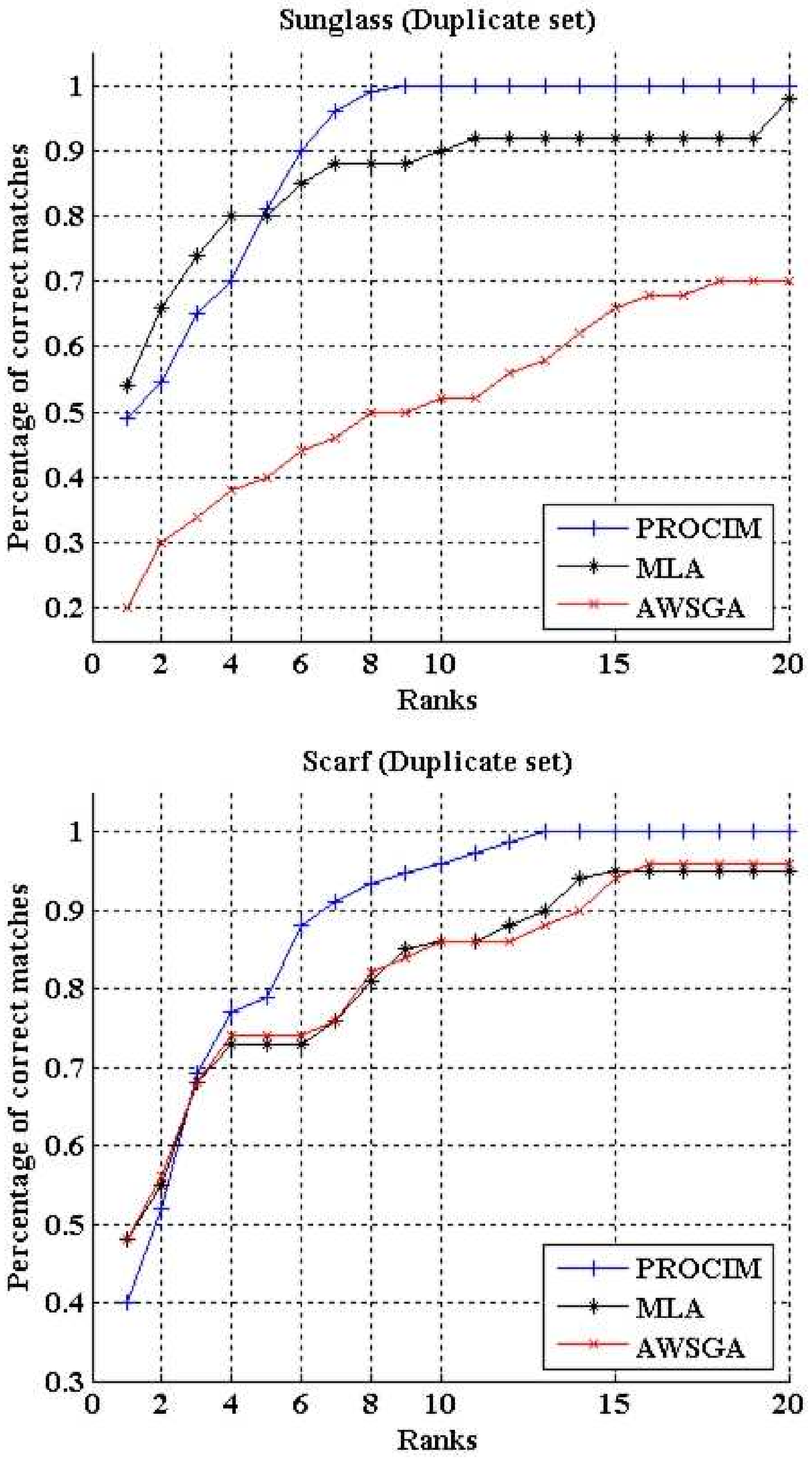


Figure 3.26: PROCIM versus MLA duplicate set

Further we have compared PROCIM with MLA against synthetic occlusions. Martinez [125] represent a whole face in terms of 6 local areas (elliptical components) and considers different occlusion sets, denoted as occ_h with $h = \{2, 3, 4\}$. However what is the psychological plausibility or atleast the scientific basis of segmenting the face image into elliptical components is not evident. Respectively the elements $2, 3, 4$ in the set indicate the increasing quantity of occlusions, that is any two, three, four out of six local areas of a given probe face have been occluded. This in turn infers that about 33%, 50% and 66% of synthetic occlusions have been simulated on the probe face images at various possible combinations. The CMC response of MLA and PROCIM for these occlusion sets are shown in Fig.3.27. The graphs show that for the first two ranks, MLA performs better than PROCIM. But gradually PROCIM advances and converges to perfect recognition (100%) even in the presence of major ($> 50\%$) occlusions within 10 to 12 ranks. MLA has the advantage of using one training sample per class but with the drawback that it does not converge within reasonable ranks. However PROCIM provides a solution within 12 ranks which is a significant improvement.

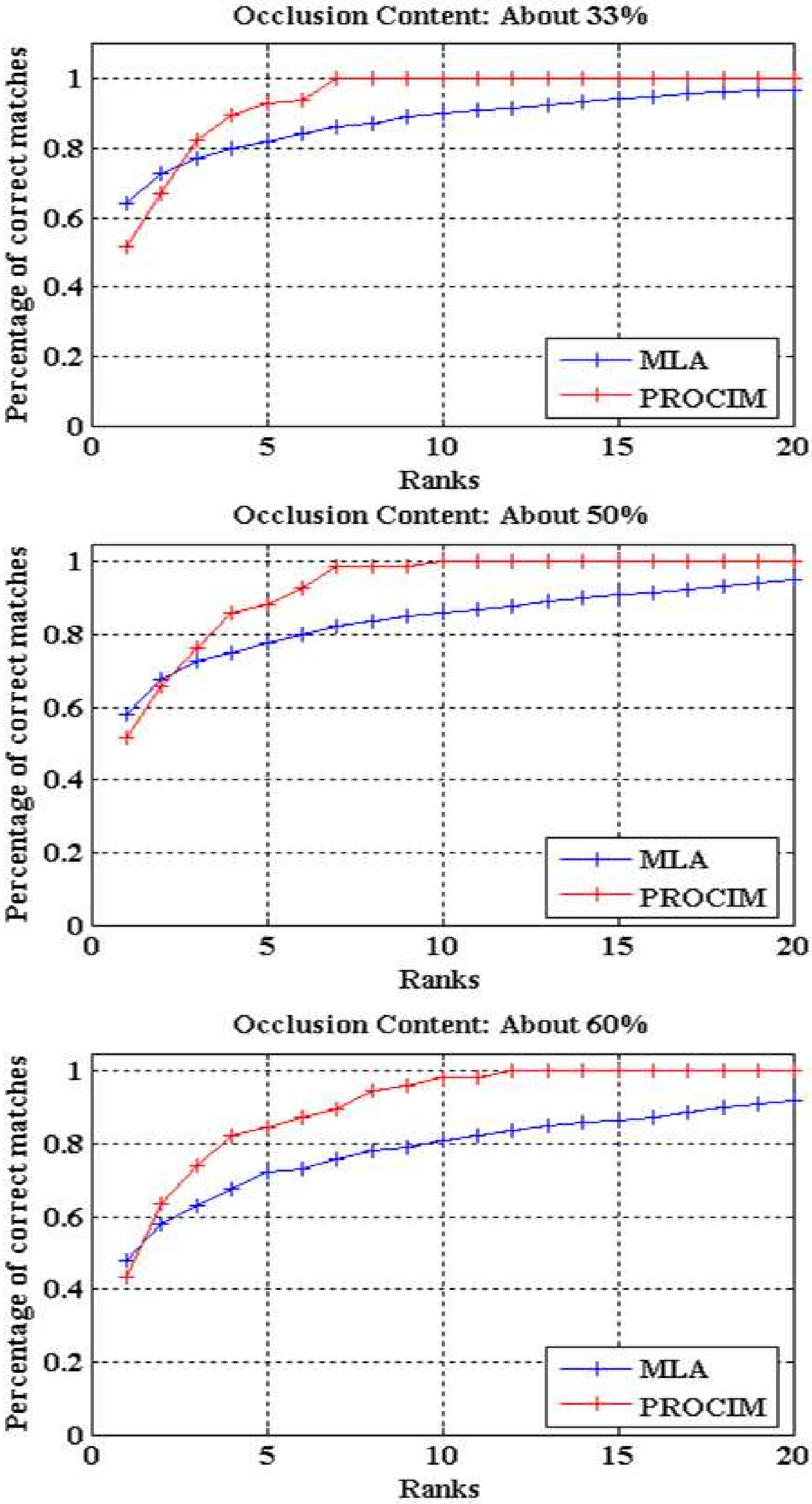


FIGURE 3.27: PROCIM versus MLA duplicate set

We subjected the FERET datasets A,B and C to CROT and their CMC response are shown in the graphs of Fig.3.28,3.29 and 3.30 respectively.

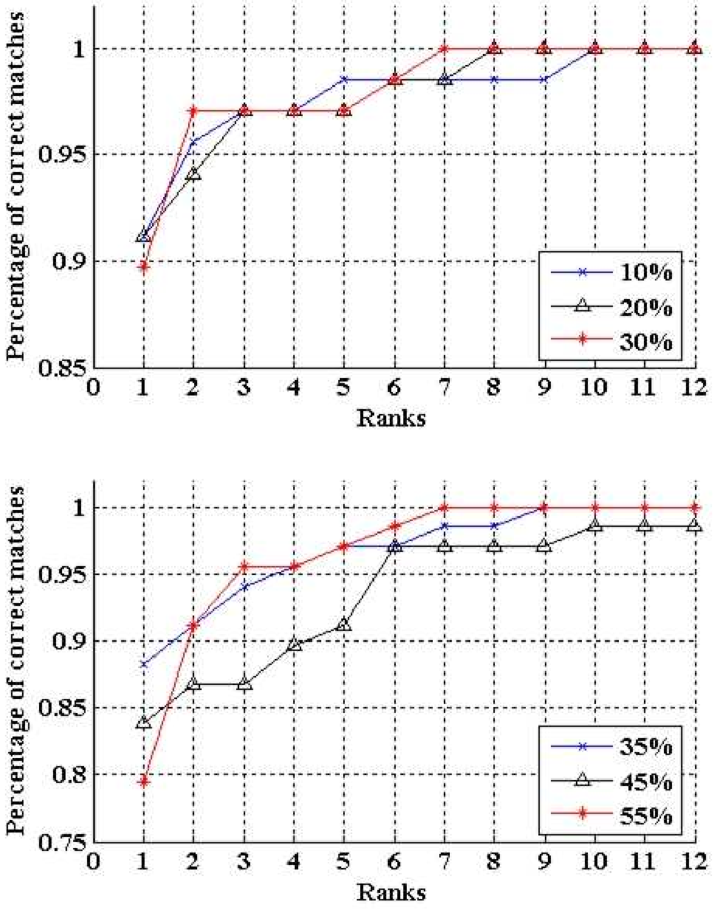


Figure 3.28: PROCIM subjected to CROT on FERET dataset A

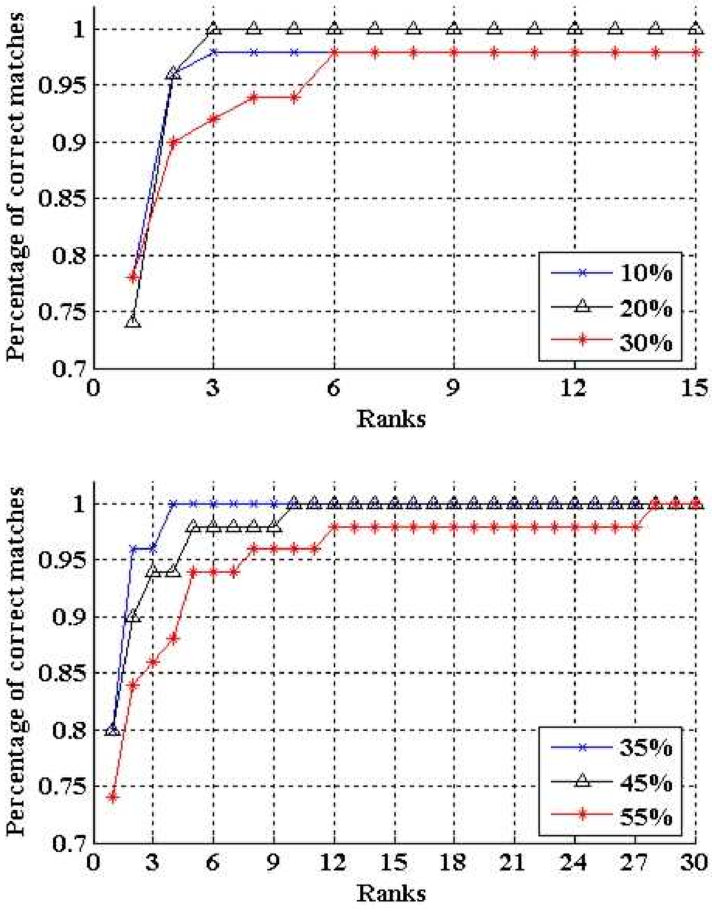


FIGURE 3.29: PROCIM subjected to CROT on FERET dataset B

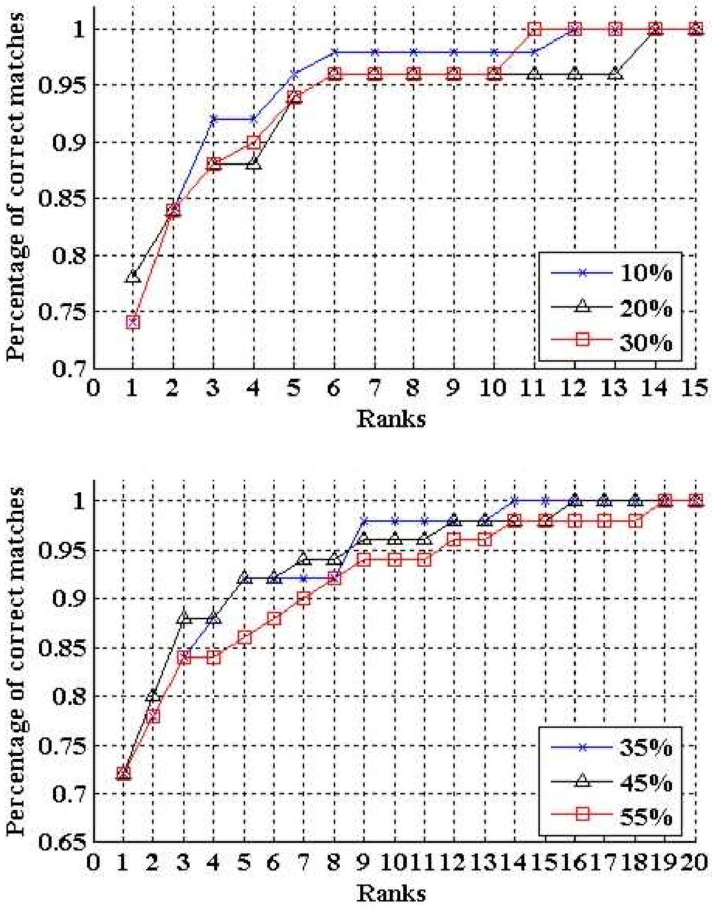


FIGURE 3.30: PROCIM subjected to CROT on FERET dataset C

We have quantified the occlusion content simulated over the probe face images in terms of percentage and classify the continuous random occlusion test (CROT) into two categories viz., I and II. I represents probe images subject to minor to moderate occlusion content comprising 10%, 20% and 30% occlusions and that of II represents major occlusion content of 35%, 45% and 55%. As the occlusions are simulated on random locations, the performance may not be directly proportional to the occlusion content. Hence in order to have a fair evaluation, the mean performance of occlusion contents (each of I and II) are reported. The overall performance for the three datasets yielded by PROCIM for I and II are 94.3% and 90.1% respectively within the top three ranks. We have compared the proposed PROCIM model with PCA against DROT which are shown by the graphs in Fig.3.31.

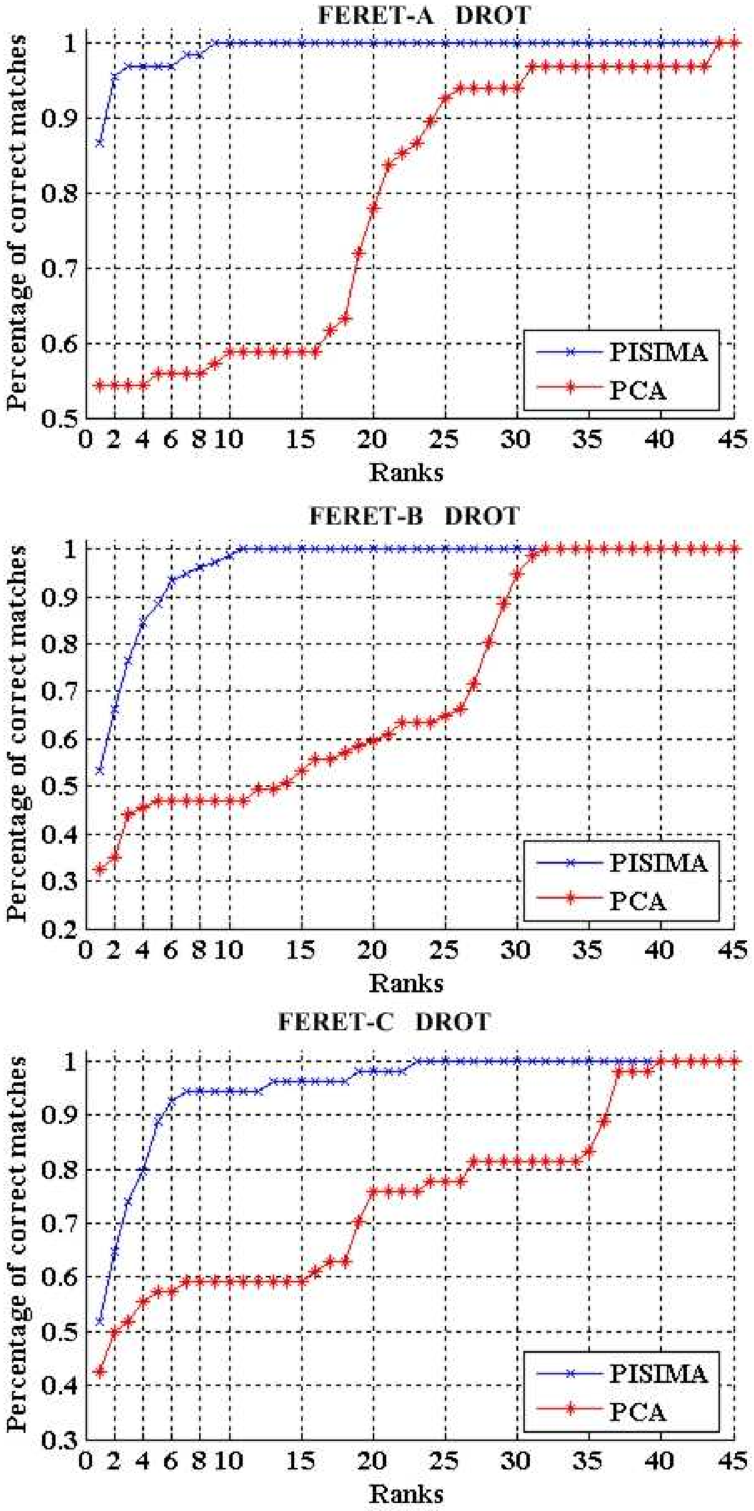


Figure 3.31: PROCIM subjected to DROT on FERET datasets A,B and C

For the discrete random occlusion test (DROT), PROCIM reports an overall 82.7% performance within the top three ranks. These results show that DROT throws a major challenge compared to CROT. The wide gap seen in the graphs between PROCIM and the conventional PCA justifies the fact that component based object models perform much better than unified object models when a high number of uncertainties including occlusions are present.

3.10 Summary

We have discovered that faces exhibit interesting similarity mappings. The proposed framework intuitively exploit these intrinsic mappings to recognize faces when they are prone to major occlusions coupled with other variations. The proposed PROCIM model encapsulates key psychophysical principles fundamental to reasoning under uncertainty, by means of statistical machine learning techniques. Compared to the state-of-the-art techniques, PROCIM reports improved recognition rates. The fact that PROCIM has the ability to converge to peak performance within a few top ranks indicates that PROCIM promises to recognize the actual suspect whose face contains major occlusions within a smaller number of ranks. If a biometric enabled security system can provide such an ability, it will give the criminal investigation team a considerable advantage, which is a significant advancement in the field of biometrics. The Discrete Random Occlusion Test (DROT) introduced in this chapter is more practicably feasible and proves to be tougher than conventional tests which simulate occlusions in terms of single continuous blocks. Hence DROT would serve as a better validation measure to evaluate future occlusion models. Further we have shown that less parameters are sufficient to build the model and hence PROCIM does not demand special computing resources.

Chapter 4

Overview of Gait Recognition

4.1 Introduction

Though the development of biometric algorithms started in the mid-1960's with work on fingerprint, face and speaker recognition, computer vision based recognition approaches to gait were first developed only in the early 1990s. If we recall the comparative analysis of biometric technologies by market share (Fig.1.2) which we presented in Section 1.1, it is clear that Gait Recognition (GR) technology is yet (even in 2009) to explicitly yield any significant commercial contribution. However, interest on GR research is driven and promoted actively by "DARPA's Human ID at a distance program". The reason is that gait as a biometric offers the unique advantage of recognizing people at a distance from low resolution videos unobtrusively, where application of other biometrics are not feasible. GR differs from gait classification which classifies human motion into categories such as walking, running and jumping. Recognizing human emotion, that is identifying whether a human is in one of the states of anger, disgust, fear, joy, sadness and surprise, using gait data is another active research area which comes under gait analysis. Though each individual is characterized by some unique walking behavior, gait is a complex spatio-temporal biometric and not very distinctive. It is gradually getting well accepted as a biometric for surveillance applications. Recent advances in computing technology, especially high-speed processors, bulk storage and memory resources, are enabling gait as a practicable biometric right now, although the idea has emerged decades ago [163].

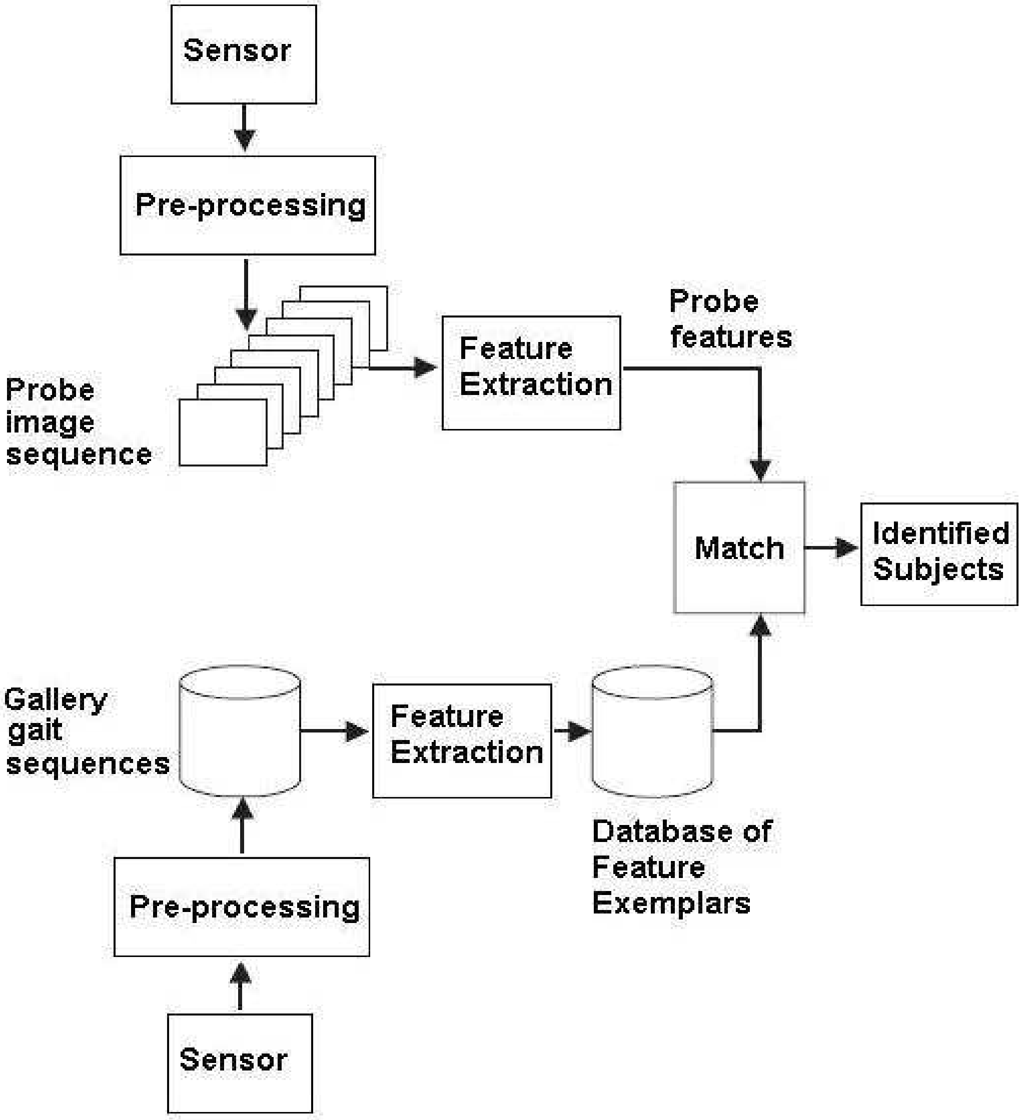


FIGURE 4.1: Flowchart showing the typical stages of a GR identification system

In the class of object recognition problems, identification is considered to be harder than verification [164]. Although gait can disclose more than identity, it is increasingly being applied to identification tasks [165]. Though there exists many GR approaches, they usually follow the stages shown in Fig.4.1. The sensor serves as an interface between the real world and the GR system. Its role is to acquire all the necessary data. The raw data acquired by the sensor need to be preprocessed. This includes removal of artifacts, background noise, minimize variations caused by illumination variation and so on. Then, most discriminative features are extracted from the preprocessed data by discarding redundant information and a database of feature exemplars, training samples, are formed. A given probe image sequence undergoes similar stages like the gallery ones until probe features are extracted. Finally, probe features are compared with gallery features using similarity measures and subjects are identified. We don't address the verification problem where a single probe is matched with a single gallery, a one-to-one match. Rather we focus on the identification problem which matches a given probe gait sequence against a set of gallery gait sequences, a one-to-many match.

4.2 Motivation from other fields

Literature reveals that several computer vision based GR techniques have derived their basis from cross disciplinary areas such as psychology, biomechanics and medical analysis. Computer vision oriented gait recognition is inspired by some historical work on human locomotion research. Borelli(1608-1679), who is regarded as the father of biomechanics, showed interests on the mechanical principles of locomotion. His study is considered as a starting point for the study of biomechanics of locomotion [20]. Later, Weber brothers (1836) investigated human gait, on aspects of both walking and running with simple instrumentation, and suggested that the lower limbs act like a pendulum. However, these awaited scientific justification. More advanced mathematical techniques and reliable instrumentation were necessary to probe into the study of locomotion. Muybridge was the first to employ photographic techniques extensively to record locomotion [166].

Literature shows that in the early 1970's medical studies have first tried to treat gait as a discriminating trait [167]. The task of classification of gait components plays a vital role in medical research in order to aid the treatment of pathologically abnormal patients. Murray et al. [3] introduced a cost effective way using reflective strips to specific anatomic landmarks of human body to generate gait patterns. Although, in the perspective of today's standards, this appeared to be a crude method, the sagittal plane joint angle measurements of normal subjects in her publications are very similar to those obtained with current technology. She compared the gait patterns of pathologically normal people with that of pathologically abnormal patients. The periodic behavior of hip motion computed from gait sequences of a typical individual is shown in Fig.4.2, where the upper and lower curves indicate standard deviation. In the first half of the gait cycle, the hip is in continuous extension as the trunk moves forward over the supporting limb. In the second half, once the weight has been passed onto the other limb, the hip flexes in preparation for the swing phase. Her study revealed that the hip (pelvic) and thorax (thigh) rotations highly varied from one subject to another. Many recent biometric research works capitalize on these findings [168–172].

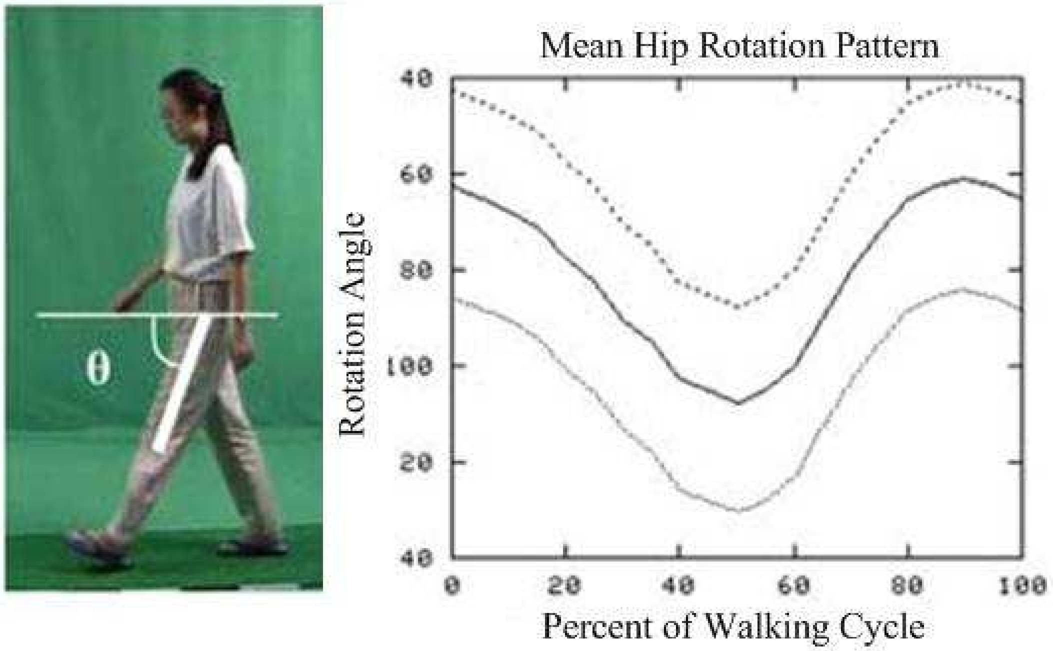


FIGURE 4.2: A typical result from early medical research ([3]) still serves as a basis for many gait based biometric research. The graph shows that hip motion within a gait period exhibits some regularity. Even recent GR techniques capitalize on this idea.

The ability of humans to recognize gaits has long been of interest to psychologists. Johansson [167] showed that humans can quickly (in less than one second) identify that a pattern of moving lights, called a moving light display (MLD), corresponds to a walking human. However, when presented with a static image from the MLD, humans are unable to recognize any structure at all. For example, without knowing that the dots in a single frame of the sequence shown in Fig.4.3 are on the joints of a walking figure, it is difficult to recognize them as such. Further it is difficult to show in a print medium, that within a fraction of a second after the dots move, one can recognize them as being from a human gait.

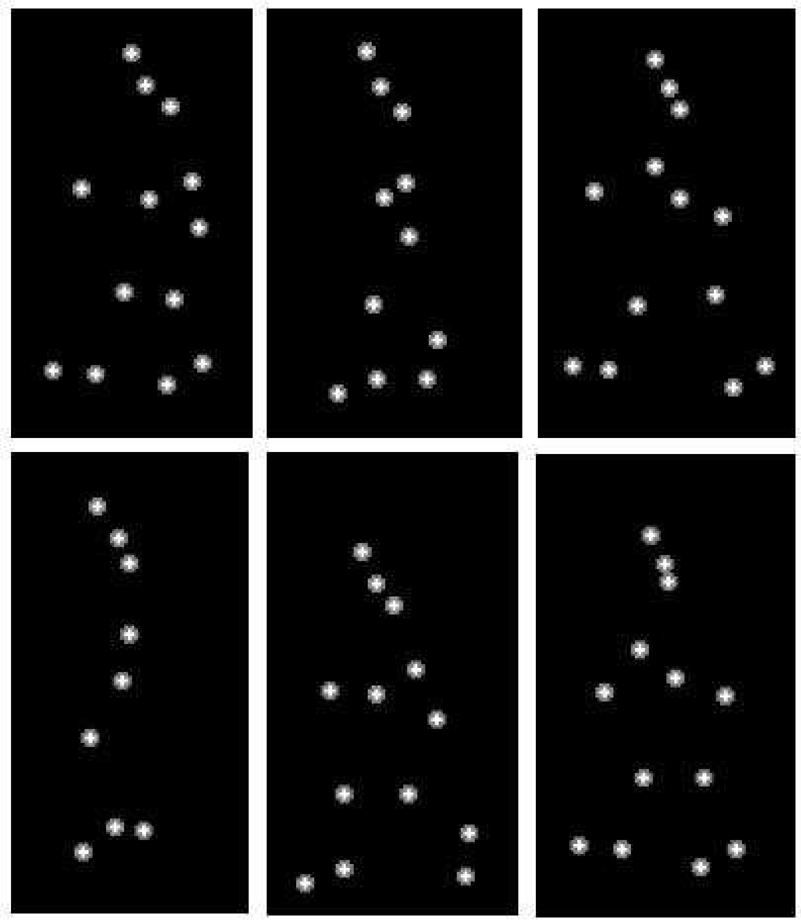


FIGURE 4.3: Frames from a moving light display of a person walking [4]. People can quickly identify that the motion is a gait from the moving sequence, but have difficulty with static frames. Decades back, the results of this psychological study revealed that people can recognize their friends from motion but motion alone is not sufficient to be a reliable form of identification.

Johansson's findings have been even referred in recent GR papers (Eg.[173, 174]) as they provide an empirical method that allows one to view motion extracted from other contextual information. Kozlowski and Cutting [175] showed that humans can recognize the gender of a walker from an MLD. However, for short exposures to the MLD (two seconds or less), humans could recognize gender at a recognition rate of only 50%. It required longer exposures, on the order of four seconds, for humans to perform a better recognition. Even then, a recognition rate of about 66% only has been recorded. Cutting and Kozlowski [4] also showed that people can recognize their friends from MLDs. The experiment involved six students who knew each other well. Experimenters recorded MLDs for the six students. Then, at a later date, the original six, plus a seventh one, who has been also a friend, tried to recognize their friends from the MLDs. This yielded a recognition rate of 38% which is significantly better than recognizing a friend at random (17%). These results concluded an important fact that people can recognize their friends from motion, but motion alone is not sufficient to be a reliable form of identification. Later work showed that point light displays aided to classify gaits as different types of motion such as jumping and dancing [176]. Later, Binham [177] showed that point light displays are sufficient for the discrimination of different types of object motion and that discrete movements of parts of the body can be perceived. As such, human vision appears adept at perceiving human motion, even when viewing a display of light points. Indeed, the redundancy involved in the light point display might provide an advantage for motion perception [178] and could even offer improved performance over video images. A recent study [179], using video rather than point light displays, has shown that humans can indeed recognize people by their gait, and learn their gait for purposes of recognition. As an outcome of this study, it has been confirmed that, even under adverse conditions, gait could still be perceived.

Boyd [180] proposed a psychologically plausible GR technique which has been derived based on principles from psychology and biomechanics. He used phase-locked loops, a technique which applied the following three important properties of human perception about gaits hypothesised by Bertenthal and Pinto [181]:

- *Frequency entrainment:* The various components of the gait must share a common frequency.
- *Phase locking:* The phase relationships among the components of the gait remain approximately constant. The lock varies for different types of locomotion such as walking versus running.
- *Physical plausibility:* The motion must be physically plausible human motion.

Boyd's interpretation is as follows. There are motions at different frequencies within a gait. However, the overall gait has a fundamental frequency that corresponds to the complete cycle. Other frequencies are multiples of the fundamental frequency. This phenomenon is termed as frequency entrainment which infers that it is not possible to walk with component motions at arbitrary frequencies. When the motions are at entrained frequencies, the phase of the motions must be locked, i.e., the timing patterns of the motions are fixed. In a typical gait, the left arm swings in phase with the right leg and opposite in phase with the left leg, a pattern that is fixed throughout the gait called phase locking. Further, Boyd gathered evidence from biomechanics literature [182–185] and hypothesized the following related to gait perception:

* People have an internal gait model that is used to synthesize their own gait. This model is a combination of a person's own kinematic structure that has an innate ability to walk and a control system that produces variations of the gait as needed.

* Humans use this internal gait model to perceive the gaits of others.

The above hypotheses suggested him a way to build a GR technique that perceive gaits by synchronizing an external stimulus with an internal gait model. The technique used arrays of phase-locked loops, called video phase-locked loops (vPLLs) to synchronize a system with the oscillations in pixel intensities that occur when viewing a gait or other oscillating stimulus. A recent GR approach proposed by Wagg and Nixon [186] derived its basis from biomechanical literature [21]. This study investigates discriminatory potential of various gait components such as hip and ankle width. The PROCIM framework proposed in this thesis too investigates discriminatory potential of various gait components, but in the context of a *computer vision* perspective. Hence we will provide relevant insight into Wagg and Nixon's work in Section 4.3.1.2 and 5.5.2.

4.3 Related Work

In order to provide a systematic overview of related work in GR, we shall first propose a flexible classification scheme and then present various research works as per this scheme. Based on a narrow scope of a selection of automatic GR approaches, Nixon et al. [171] have figured out a taxonomy of GR models as shown in Fig.4.4. A model-based analysis usually involves fitting a model representing various aspects of the human anatomy to the video data and then extracting and analyzing its parameters (Eg.[186, 187]). On the other hand, a model-free approach utilizes the description of instantaneous motion from moving shape or integrates shape and motion within the description (Eg. [188, 189]). Nixon et al. have shown that it is hard to define boundaries between these two categories. For example the human intuition based approach proposed by Boyd [180] attempts to bridge the gap between the model-free and model-based domains. As Boyd's approach straddles the boundary between the two domains it does not fit well into the scheme shown in Fig.4.4.

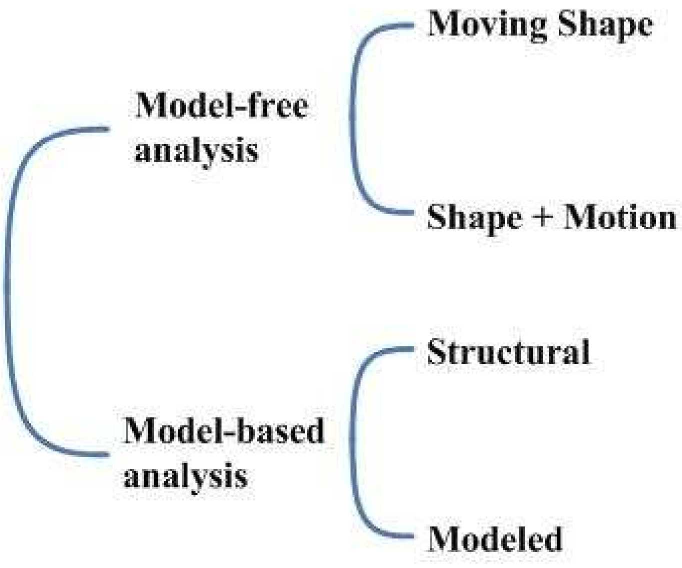


FIGURE 4.4: Nixon et al.'s taxonomy based on selected automatic GR approaches. The authors state that this classification scheme is unclear as a number of approaches straddle the boundaries.

However, on the basis of data acquisition, we propose a general classification scheme as shown in Fig.4.5 which is a hybrid of classification schemes proposed by Liu et al. [9] and Gafurov et al. [168]. This scheme aims to encompass a wide range of approaches and hence has a broader scope than the scheme proposed by Nixon et al. [171]. We will briefly present an overview of GR algorithms based on this taxonomy as follows:

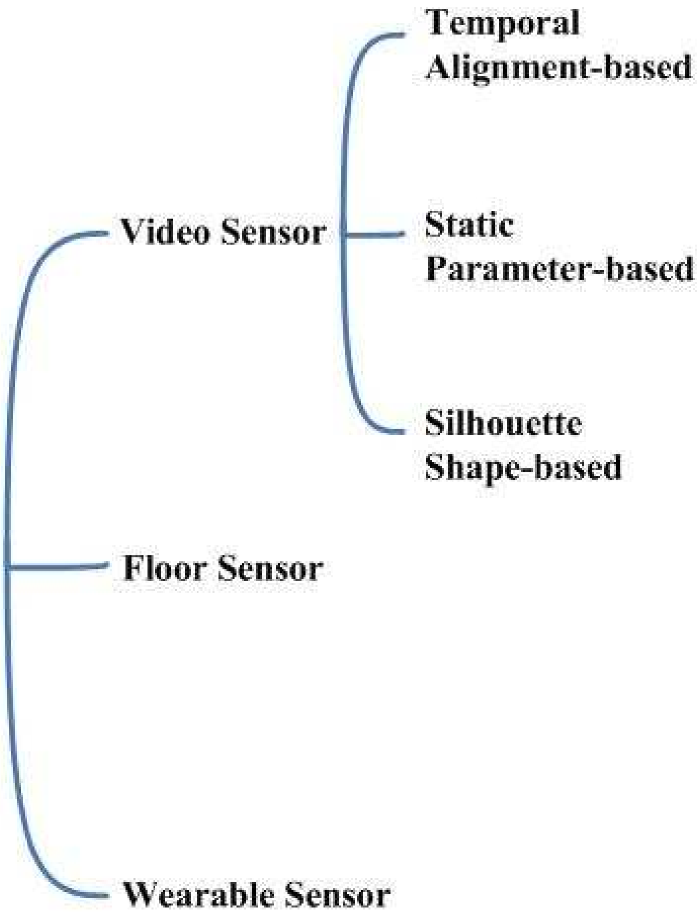


FIGURE 4.5: Proposed classification scheme of GR approaches based on acquisition of gait data. This scheme has a broader scope when compared to Nixon et al.'s scheme shown in Fig.4.4

4.3.1 Video Sensor(VS)-based GR approaches

VS-based GR can be further classified into three sub-categories namely temporal alignment-based, static parameter-based and silhouette shape-based [9].

4.3.1.1 Temporal alignment-based approaches

This category of approaches considers both shape and dynamics and treats gait sequences as time series-based patterns. Potential sources for gait biometrics can be seen to derive from two aspects viz,, shape and dynamics. Shape refers to the configuration or shape of the people as they perform different gait phases. On the other hand dynamics refers to the rate of transition between these phases and is usually the aspect one refers to gait in traditional problem contexts, such as biomechanics or human motion recognition.

A classical example which falls under this category is the *population Hidden Markov Model (pHMM)* proposed by Liu and Sarkar [9]. The proposed GR algorithm attempts to compensate the uncertainty encountered by the model, caused as a result of factors such as varying walking speed, noise and so on by normalizing the gait dynamics based on a population-based generic walking model. For each gait sequence, Viterbi decoding of the gait dynamics is applied to arrive at normalized gait cycles of fixed length. Each gait cycle is chosen to begin at the right heel strike phase of the walking cycle through to the next right heel strike. The states of the pHMM is represented by gait stances over one gait cycle and the silhouettes of the corresponding gait stances are considered as observations of the model. The model is trained on a set of manually created silhouettes and the exemplar sets are initialized by equally partitioning the frames in one gait cycle into N_s number of segments. Formally the pHMM is specified by the possible states, $q_t \in 1, \cdots, N_s$, which basically represent gait stances and the triple parameters, $\lambda = (A, B, \pi)$, which respectively represent the state transition matrix, an observation model and priors. For a given input silhouette frame, f_t, the number of observation variables were in proportion to the number of exemplars. The observation model comprises a model for each state,

$$B = \{b_j(f_t) \mid j = 1, \cdots, N_s\}, \tag{4.1}$$

where

$$b_j(f_t) = P(f_t \mid q_t = j) \tag{4.2}$$

is the conditional probability of the observed silhouette, f_t, at time t given that the state at time t is j. The observation model is chosen to be exponential in terms of the observation variable

$$b_j(f_t) = \frac{1}{\mu_j} e^{\frac{-D(f_t, E_j)}{\mu_j}}, \tag{4.3}$$

where D is the Tanimoto distance between any given silhouette, f_t, E_j is the mean of the state exemplars and

$$\mu_j = \frac{\sum_{f_i \in E_j} D(f_i, E_j)}{\mid E_j \mid} \tag{4.4}$$

The similarities between any two gait sequences is computed by applying a distance metric between the two corresponding dynamics-normalized gait cycles. The distance metrics computed between an observed silhouette and the silhouettes in the exemplar set serve as observation variables. The distance between two vertically scaled and horizontally aligned silhouettes, f_i and f_j, is defined as:

$$S(f_i, f_j) = \frac{f_i^T f_j}{f_i^T f_i + f_j^T f_j - f_i^T f_j} \tag{4.5}$$

Distances between two silhouettes from the same generic gait stance is computed in the linear discriminant analysis space so as to maximize the discrimination between persons, while minimizing the variations of the same subject under different conditions. This aided the approach to handle variations in silhouette shape, due to dilations and erosions that could occur with changing imaging conditions. The authors conclude that shape is more significant than dynamics (kinematics) for person identification as dynamics is vulnerable to uncertainty factors. The authors stress that due to high intrasubject variability, dynamics might not be a stable source for biometric information.

Veeraraghavan and Chowdhury [10] have compared the role of shape and kinematics in automated gait-based person authentication. They have proposed a Dynamic Time Warping (DTW) algorithm. The objective is to learn the dynamics of shape changes in a gait sequence using the distance measures between shape sequences. They have applied autoregressive moving average models. The DTW algorithm which is based on dynamic programming computes the best nonlinear time normalization of the test sequence in order to match the template sequence by performing a search over the space of all allowed time normalizations. The algorithm derives geometric information of the walking person from several landmark points which are manually marked on the gait video. The space of all time normalizations allowed is cleverly constructed using certain temporal consistency constraints which are specified as follows:

- The beginning and the end of each sequence is rigidly fixed. For example, if the template sequence is of length N and the test sequence is of length M, then only time normalizations that map the first frame of the template to the first frame of the test sequence and also map the Nth frame of the template sequence to the Mth frame of the test sequence are allowed.
- The warping function (mapping function between the test sequence time to the template sequence time) should be monotonically increasing. In other words, the sequence of events in both the template and the test sequences should be the same.
- The warping function should be continuous.

The distance $D(A(t), B(t))$ between two shape sequences $A(t)$ and $B(t)$, is defined as

$$D(A(t), B(t)) = DTW(A(t), B(t)) + DTW(B(t), A(t)), \tag{4.6}$$

where

$$DTW((A(t), B(t)) = 1/T \sum_{t=1}^{T} d(A(f(t)), B(g(t))), \tag{4.7}$$

where f and g are the optimal warping functions, d is a distance function based on *Procrustes shape analysis*[190, 191]. The preshape vector extracted by the method lies on a spherical manifold. Therefore, a concept of distance between two shapes must include the non-Euclidean nature of the shape space. Keeping this in view, a Procrustes distance metric has been applied which serves as a similarity measure between sequences of deforming shapes. Ultimately, the DTW function aims to normalize shape sequences based on a fixed time frame of T units. Such a distance between shape sequences is commutative. The isolation property, i.e., $D(A(t), B(t)) = 0$ iff $A(t) = B(t)$, is enforced by penalizing all non-diagonal transitions in the local error metric. The results of this analysis suggests that kinematics helps to boost recognition performance but it is not sufficient as a stand-alone feature for person identification; Human body shape-based algorithms perform better than purely kinematics-based algorithms. The authors conclude that shape is more significant for person identification than kinematics.

The alignment process which plays a key role in temporal alignment-based approaches has been modeled using several techniques such as simple temporal correlation [12, 189], dynamic time warping [10], hidden Markov models [10, 188], phase locked-loops [180] and Fourier analysis [192].

4.3.1.2 Static parameter-based approaches

This second category of approaches characterize the human motion based on parameters such as stride length, cadence and stride speed. Sometimes static body parameters, such as the ratio of sizes of various body parts are considered in conjunction with these parameters. A typical GR approach which falls under this category is the one proposed by BenAbdelkader et al. [193]. The proposed method aims to automatically estimate the spatio-temporal parameters of gait (stride length and cadence) of a walking person from video. Stride and cadence estimates computed based on body height, weight and gender are used as biometrics for the problem of human identification and verification. Cadence is estimated using the periodicity of a walking person. Using a calibrated camera system, the stride length is estimated by first tracking the person and estimating their distance traveled over a period of time. By counting the number of steps (using periodicity) and assuming that subjects are walking at constant-velocity, strides are estimated. The proposed technique [193] makes the following assumptions:

- People walk on a known plane with constant velocity (i.e. in both speed and direction) for about 10-15 seconds (i.e. the time for 20-30 steps).
- The camera is calibrated with respect to the ground plane.
- The frame rate is greater than twice the walking frequency.

Initially the walking subject in each frame is tracked, binary silhouettes are extracted, and the subject's 2D position in the image is computed. Once a person has been tracked for a certain number of frames, the gait period T, in frames per cycle, the distance W traveled, in meters, are estimated. Then the cadence C, in steps per minute and stride length, in meters, are computed using:

$$C = \frac{120 \times F_s}{T}, \tag{4.8}$$

$$L = \frac{W}{n/T}, \tag{4.9}$$

where n is the number of frames and F_s is the frame rate (in frames per second), n/T is the number of gait cycles traveled over the n frames. Assuming that the subject is walking in a straight line, the total distance traveled, which is simply the distance between the first and last $3D$ positions on the ground plane is given by

$$W = \parallel P_n - P_1 \parallel \tag{4.10}$$

The subject's $3D$ position, (X_g, Y_g, Z_g), is computed from the $2D$ position image, (x_g, y_g) as follows. Given the camera matrices, K, E and the parametric equation of the plane of motion,

$$P : aX + bX + cZ + d = 0, \tag{4.11}$$

and assuming perspective projection a linear system of equation is formulated using

$$\begin{pmatrix} k_{11} & 0 & -x_g + k_{13} \\ 0 & k_{22} & -y_g + k_{23} \\ \hat{a} & \hat{b} & \hat{c} \end{pmatrix} E \begin{pmatrix} X_g \\ Y_g \\ Z_g \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -d \end{pmatrix}, \tag{4.12}$$

where

$$(\hat{a}, \hat{b}, \hat{c}, \hat{d}) = (a, b, c, d)E^{-1} \tag{4.13}$$

and k_{ij} is the $(i, j)^{th}$ element of K. One of the main problems with the above formulation is that the equation 4.12 does not have a unique solution if the person is walking directly towards or away from the camera (i.e. along the optical axis). With a small database of 17 people and 8 samples of each, the authors report that a person is verified with an Equal Error Rate (EER) of 11%, and correctly identified with a probability of 40%. Static parameter-based based approaches have not reported good performance on common databases, partly due to their need for 3D calibration information [9].

Collins et al. of Carnegie Mellon University (CMU), presented a simple GR algorithm for human identification from body shape and gait. This CMU algorithm is based on matching 2D silhouettes extracted from key frames across a gait cycle sequence. These key frames are compared to training frames using normalized correlation, and subject classification is performed by nearest neighbor matching among correlation scores. The approach implicitly captures biometric shape cues such as body height, width, and body-part proportions, as well as gait cues such as stride length [11].

Wagg and Nixon [186] proposed a GR approach, guided by biomechanical analysis [21], which explicitly uses structural body components as part of generating shape models consistent with normal human body proportions. Based on the results from a statistical analysis of the extracted gait parameters, this study suggests that recognition capability, is primarily gained from cadence and from static shape parameters, although gait is the cue by which these parameters are derived. Images acquired from gait video sequences are pre-processed using a Gaussian averaging filter for noise suppression, followed by Sobel edge detection and background subtraction (the background is computed as the temporal median of neighbouring frames). This removes all static objects, leaving only edges belonging to moving objects. Then, an estimate of human shape is derived by shifting and accumulating the edge images according to

$$A_v(i,j) = \sum_{t=0}^{N-1} E_t\left(i + v\left(\frac{N}{2} - t\right), j - dy_t\right) \tag{4.14}$$

where A_v is the accumulation for velocity v (in pixels per frame), E_t is the edge strength image at frame t, i and j are coordinate indices, N is the number of frames in the gait sequence and dy_t is the y-displacement of the subject from their center of oscillation at frame t. Each moving object in the scene will appear as a peak in a plot of maximal accumulation intensity against velocity. If the subject is the most significant moving object in the scene (in terms of edge strength and visibility), their velocity can be inferred by selecting the highest peak in this plot. The subject is then extracted by matching a coarse person-shaped template which is constructed from data, gathered from biomechanical source [21], scaled to subject's height. Basic geometrical shapes viz., a trapezium, line segments and rectangles are employed to capture body components such as legs, foot, torso and the head. The segmentation technique is constrained by mean anatomical proportions. Gait frequency is determined by finding the frequency and phase that minimizes the error function computed using

$$X_s = \sum_{t=0}^{N-1} \left(S_t - A_s sin2\left(\omega_i t + \phi_j + \frac{\pi}{11}\right)\right)^2 \tag{4.15}$$

where X_s is the energy function to be minimized, N is the number of frames, S_t is the normalized signal magnitude at frame t, A_s is the sinusoid amplitude,

ω_i and ϕ_j are the proposed gait frequency and phase. The offset phase determined empirically has been used to align the sinusoids. The initial estimates of the shape parameters derived from hip, knee and angle are computed as fixed proportions of the subject's height. Some subjects might wear loose clothings such as baggy trousers or skirts. This might hinder the accuracy of shape parameters. By applying Hough transforms for each frame within the upper and lower leg regions, improved estimates were obtained in order to mitigate the uncertainty factors. However, a significant reduction in discriminatory capability in features, extracted from the outdoor dataset compared to those from the indoor dataset has been observed. Further an overall estimate of leg width is computed as a mean of the best line parameters from each frame, weighted by accumulation intensity. The experimental results of this study reported in terms of cumulative match characteristics exhibit Correct Classification Rates (CCR) of 84% and 64%, for the indoor and outdoor datasets, respectively. The result of this study shows an important fact that recognizing gaits under outdoor settings is more difficult than that of indoor settings.

4.3.1.3 Silhouette-based approaches

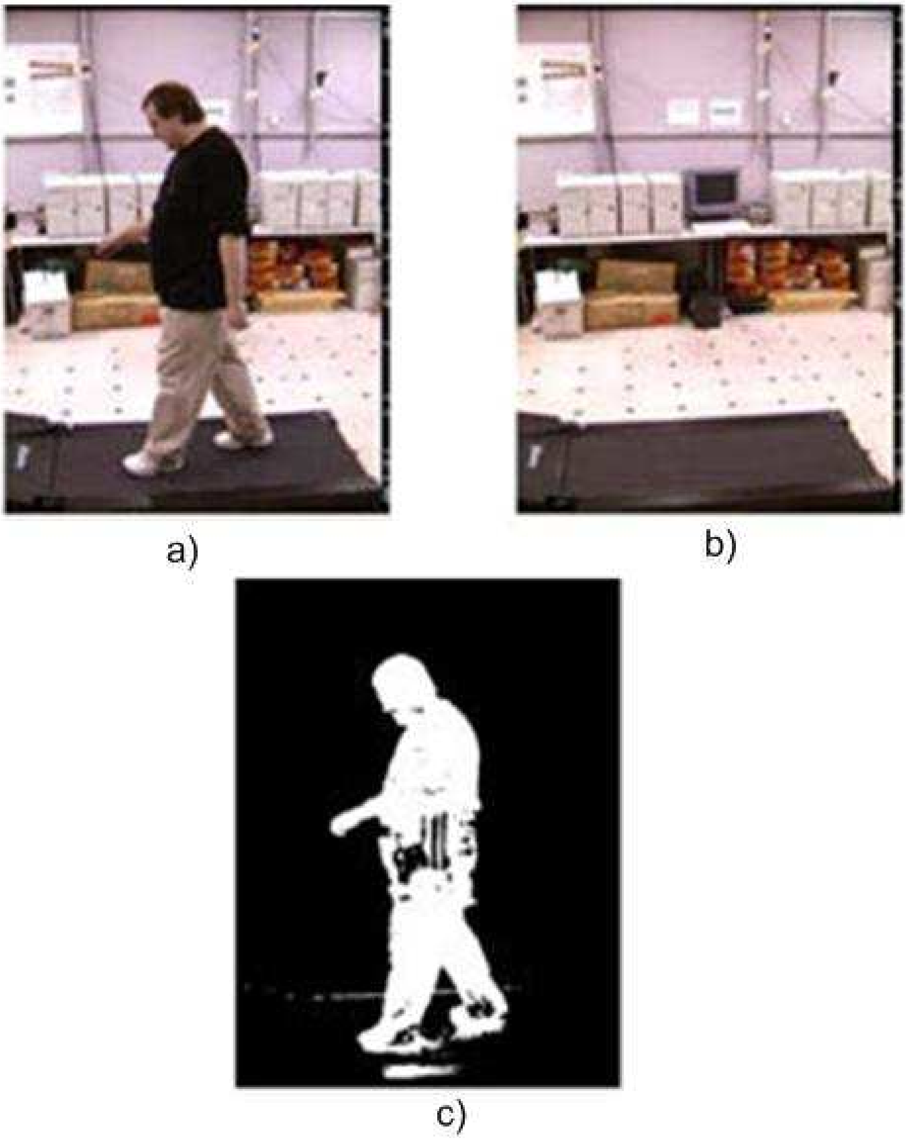


Figure 4.6: The key stages of a background subtraction technique being illustrated using the CMU Mobo data set [5] : a) A scene of a gait video frame b) The built background model of the scene c) The final binary silhouette extracted by background subtraction.

Intuitively the silhouette, which represents the binary map of walking humans, forms a robust feature to represent gait. Silhouettes representing human walks are extracted from gait videos using a procedure called background subtraction, the process of segmenting foreground pixels representing the walking human, from the background of the image sequence. A given video frame, its background model and the finally segmented binary silhoutte image are shown in Fig. 4.6. Standard gait dataset providers such as, the Chinese academy of sciences [194], offer a silhouette database. Hence silhouettes form the core input of many GR algorithms (Eg. [195, 196]). Silhouette-based gait recognition techniques are gaining much interest among current gait recognition researchers [165, 195, 197]. The reason is that they do not need any further information such as color, texture or gray-scale metrics and they capture the motion of most of the body parts [188]. Recent studies [11, 198] have shown that silhouette shape has equal, if not more, recognition potential than gait kinematics as referred by [164].

The baseline GR algorithm proposed by Sarkar et al. [12] (University of South Florida) extracts silhouettes using an expectation maximization (EM) procedure and performs recognition by temporal correlation of silhouettes. For a given gait video frame, Sarkar et al.[12] define the silhouette as the region of pixels from a person. Initially bounding boxes around the moving subject in each frame of the gait sequence are manually defined. Then silhouettes are extracted from the bounding boxes as follows. Initially a gait sequence is parsed and its background statistics of the RGB values at each image location, (x, y), using pixel values outside the manually defined bouding boxes are computed. For pixels within the bounding box of each frame, the Mahalanobis distance for the pixel value from the estimated mean background value is computed. At each pixel, indexed by k, a two-class Gaussian mixture model, $\{Foreground = \omega_1, Background = \omega_2\}$, is imposed based on observing the Mahalanobis distance, d_k as,

$$P(d_k) = \sum_{i=1}^{2} P(\omega_i) p(d_k \mid \omega_i, \mu_i, \sigma_i) \tag{4.16}$$

where the class likelihood is computed using

$$p(d_k \mid \omega_i, \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{\frac{-(d_k - \mu_i)^2}{2\sigma_i^2}} \tag{4.17}$$

The posterior estimate $P(\omega_1 | d_k)$ is estimated using the EM procedure. The EM process is initialized by choosing class posterior labels based on the observed distance; the larger the Mahalanobis distance of a pixel from the mean, the greater is the initial posterior probability of the pixel being away from the foreground which is formulated as,

$$P^{(0)}(\omega_1 \mid d_k) = min(1.0, d_k/255) \tag{4.18}$$

$$P^{(0)}(\omega_2 \mid d_k) = 1 - P^{(0)}(\omega_1 \mid d_k) \tag{4.19}$$

Though most of the silhouettes extracted using the above formulation are in good quality, some of the silhouettes are noisy due to segmentation errors. These segmentation errors occur due to the following factors:

- Shadows formed as a result of illumination variation encountered in typical unconstrained scenarios.
- Some parts of the subject are misclassified as background (Eg. hair color could merge with the background color)
- Moving objects in the background (Eg. moving grass leaves when the subject is walking on grass)
- Lingering compressing artifacts near the boundary regions of the subject

Once silhouettes are extracted, they are normalized in terms of gait cycles, which is a crucial step in GR. This is performed by detecting the gait period using the following strategy. The number of foreground pixels in the silhouette in each frame is computed. This number will reach a maximum when the two legs are farthest apart (full stride stance) and drop to a minimum when the legs overlap (heels together stance). As a strategy to increase the sensitivity, the bottom half of the silhouette contributing to the leg portion, is taken into account. Then gait cycles are formed with the set of silhouettes between two consecutive strides.

Gaits are then classified based on comparing similarity scores between all the gallery and probe gait sequences. Similarity scores are computed by spatial-temporal correlation. The similarity between two silhouette frames, S_i, S_j, is computed as the ratio of the number of pixels, Num, in their intersection to their union. This measure is called as Tanimoto similarity measure, $Tanimoto$, which is given by:

$$Tanimoto(S_i, S_j) = \frac{Num(S(i) \cap S(j))}{Num(S(i) \cup S(j))} \tag{4.20}$$

The proposed baseline algorithm yielded a recognition rate of 72% when tested on the CMU Mobo dataset [5]. Notably the CMU Mobo dataset is popular as it has gait sequences recorded with a speed variation. Interestingly, the subjects were asked to walk over a treadmill which enables GR researchers to perform speed controlled gait recognition studies. In Chapter 6 of this thesis, we will present our results relevant to the GR speed challenge.

Wang and Tan et al. [189], Chinese Academy of Sciences (CAS), proposed a GR method based on spatiotemporal silhouette analysis. This is another typical study which further reinforces that, gait recognition via analysing, "How does silhouette shape of a walking individual change over time?", is a sensible paradigm. The authors perceived gait motion as a sequence of static body poses; distinguishable signatures with respect to those static body poses are extracted in terms of contours of silhouette sequences and finally recognition has been performed by considering temporal variations of those observations. Instead of considering full silhouette images, the authors analyze silhouette contours, that is the outer boundary of silhouettes and further convert them to associated 1D signals. Silhouettes are extracted from gait video using a background subtraction procedure, which applies the Least Median of Squares method(LMedS) [199], to construct the background from a small portion of image sequences, including moving objects as follows. If I represent a gait sequence including N images, then the resulting background b_{xy} is computed by

$$b_{xy} = \arg\min_{p} med_t(I^t_{xy} - p)^2, \tag{4.21}$$

where p is the background brightness value to be determined for the pixel location

(x, y), med represents the median value, and t represents the frame index ranging within $1 - N$. In image processing, the change in brightness is usually obtained through differencing between the background and current image. In the case of low contrast images, the selection of a suitable threshold for binarilization is difficult as most of the moving objects might be missed out. This is because the brightness change is too low to distinguish regions of moving objects from noise. The following extraction function is used to solve this issue to indirectly perform differencing

$$f(a,b) = 1 - \frac{2\sqrt{(a+1)(b+1)}}{(a+1)+(b+1)} . \frac{2\sqrt{(256-a)(256-b)}}{(256-a)+(256-b)}, \tag{4.22}$$

where $a(x, y)$ and $b(x, y)$ are the brightness of current image and the background at the pixel position (x, y), respectively, $0 \leq a(x, y), b(x, y) \leq 255, 0 \leq f(a, b) < 1$. The objective of this function is to detect the change sensitivity of the difference value according to the brightness level of each pixel in the background image. For each image I_{xy}, the distribution of the above extraction function $f(a(x, y), b(x, y))$ over x and y can be easily obtained. Then, the moving pixels can be extracted by comparing such a distribution against a threshold value. After the moving silhouette of a walking figure has been tracked, its outer contour is obtained using a border detection algorithm and the centroid of the silhouette (x_c, y_c) is determined. By choosing the centroid as a reference origin, the outer contour is unwrapped counterclockwise to turn it into a distance signal

$$S = d_1, d_2, \cdots, d_i, \cdots, d_{N_b}, \tag{4.23}$$

that is composed of all distances d_i, for N_b number of boundary pixels, between each boundary pixel (x_i, y_i) and the centroid, where d_i is computed using

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}. \tag{4.24}$$

Thus the original $2D$ silhouette shape has been compactly transformed to a signal in the $1D$ space. Finally by applying PCA, predominant components of gait signatures are computed and gaits are classified using the standard lower-dimensional eigenspace technique. Though a reasonable recognition rate of 75% has been demonstrated on a small indigenous database of 20 subjects, the authors report only a mean recognition rate of 39.6% with respect to the HumanID gait challenge experiments on the USF dataset [12]. Further, the method highly relies on the quality of silhouettes to perform a contour based reasoning, but in reality the silhouettes extracted are noisy, as reported by many GR approaches (Eg. [12, 200]). Hence the approach yields an inferior recognition rate.

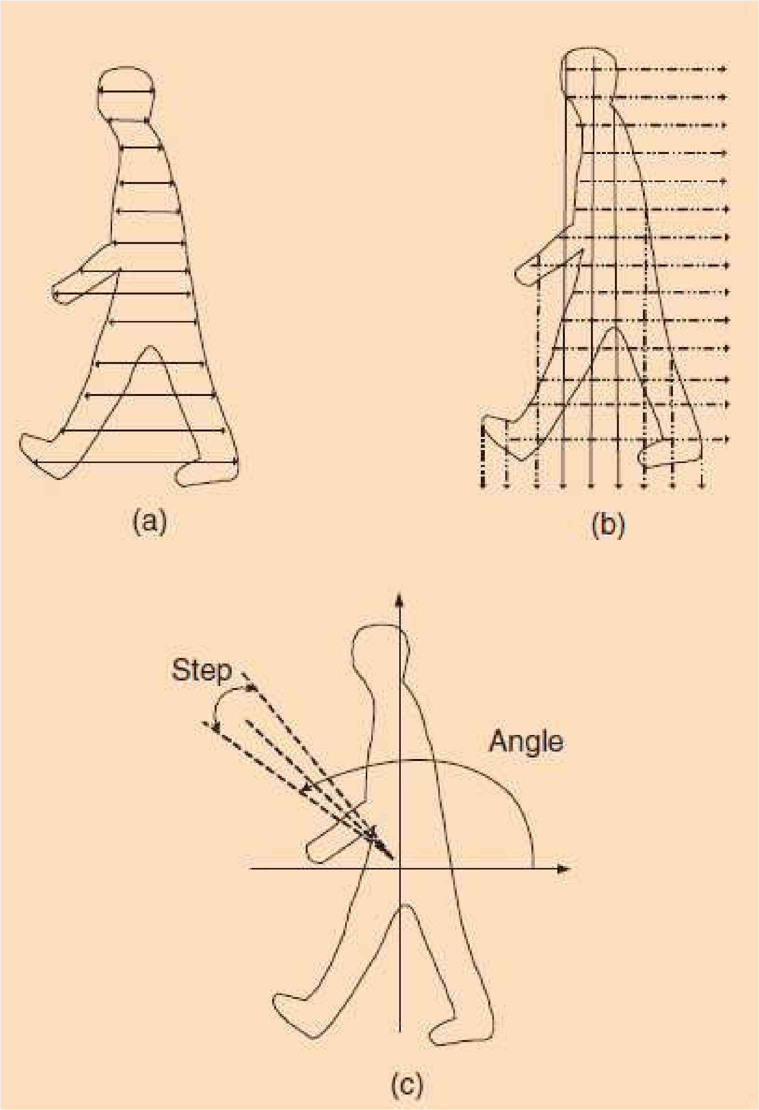


FIGURE 4.7: Typical approaches that use silhouette based features; a) An approach [6] that uses width of the silhouette b) A typical frieze pattern [7] based gait representation that maps a video sequence of silhouettes into a pair of two-dimensional spatio-temporal patterns b) Application of angular transform on binary silhouettes [8].

Kale et al. [6] suggested a GR approach where the information gained from width of silhouettes were used to discriminate gaits. The width $w(i)$ of a silhouette is defined as the horizontal distance between the leftmost and rightmost foreground pixels in each row i of the silhouette, as shown in Fig.4.7(a). Although the calculation of width signals imposes minimal processing load on a gait system, algorithms that use this feature are vulnerable to spurious pixels, that often render the identification of the leftmost and rightmost pixels inaccurate. For this reason, the authors in [6] proposed a postprocessing technique to smooth and denoise the feature vectors, prior to their deployment in gait recognition. Interestingly, the width coefficients that exhibit the greatest variance are the coefficients derived from the leg and arm area. It has been noticed in this study that shadows result in inaccurate computation of the width coefficients in the feet area.

To address GR, Liu et al. [7] proposed "*frieze patterns*", a representation that maps a video sequence of silhouettes into a pair of two-dimensional spatio-temporal patterns that are periodic along the time axis (ps. see Fig.4.7(b)). Spatio-temporal gait representations are generated by projecting the body silhouette along its columns and rows, then stacking these 1D projections over time to form 2D patterns that are periodic along the time dimension. Such a 2D pattern that repeats along one dimension is defined as a frieze pattern. Several binary silhouettes that represent a gait sequence are denoted as

$$s[i,j], \quad i = 0, \cdots, M-1, \quad j = 0, \cdots, N-1, \tag{4.25}$$

where M, N represent the number of rows and columns of the silhouette, respectively. Conversely

$$s[i,j] = \begin{cases} 1 & \text{if (i,j) belongs to the foreground} \\ 0 & \text{otherwise} \end{cases} \tag{4.26}$$

Using the above definitions, the horizontal and vertical projections of silhouettes are expressed as

$$p_h[i] = \sum_{j=0}^{N-1} s[i,j], \quad i = 0, \cdots, M-1 \tag{4.27}$$

$$p_v[j] = \sum_{i=0}^{M-1} s[i,j], \quad j = 0, \cdots, N-1 \tag{4.28}$$

The efficiency of this feature is based on the fact that it is sensitive to silhouette deformations, since all pixel movements are reflected in the horizontal or vertical projection as shown in Fig. 4.7(b). Although this feature is similar to the width of the silhouette (note the similarity between the width vector and the horizontal projection vector), it is more robust to spurious pixels. An additional advantage is that it is fast and hence can be computed in real time. With the help of a walking humanoid avatar, the authors explored the variation in gait frieze patterns with respect to the viewing angle. It has been found that the frieze groups of the gait patterns and their canonical tiles could be applied to estimate viewing direction as well. However, an important consideration here is that, the silhouettes must be centered, prior to the computation of the feature, since misplaced silhouettes will result in shifted projections.

An angular transform of the silhouette has been proposed by Boulgouris et al. [8] where the transform divides the silhouettes into angular sectors and computes the average distance between foreground pixels and the center (i_c, j_c) of the silhouette as shown in Fig. 4.7(c). This transform is computed using

$$A(\theta) = \frac{1}{N_\theta} \sum_{(i,j)\in F_\theta} s[i,j]\sqrt{(i-i_c)^2 + (j-j_c)^2}, \tag{4.29}$$

where θ is an angle, F_θ is the set of the pixels in the circular sector $[\theta-(\Delta\theta/2), \theta+(\Delta\theta/2)]$, and N_θ is the cardinality of F_θ. The transform coefficients were shown to be a linear function of the silhouette contour. This feature is, in general, robust since it obviates the need for detection of contour pixels. Each human silhouette in a gait sequence is transformed into a low dimensional feature vector consisting of average pixel distances from the center of the silhouette. The sequence of feature vectors corresponding to a gait sequence is used for identification based on a minimum-distance criterion between test and reference sequences.

Lee and Grimson [192] proposed a component based GR approach, where the silhouettes of gait sequences are subdivided into seven regions, fitted into ellipses and a set of moment related features are computed. The authors report improved recognition and gender classification rates using a small dataset, captured at indoors. Bauckhage et al. [165] proposed a method to establish homeomorphisms between 2D lattices and binary silhouettes. This method provides a robust vector space embedding of segmented body silhouettes. Feature vectors obtained from this scheme show improved detection of abnormal gait. Li et al. [174] proposed a component based approach by segmenting silhouettes into seven components, namely head, arm, trunk, thigh, front-leg, back-leg and feet. The effectiveness of these components for gait recognition and gender recognition has been analyzed. The approach relies on manually selected control points. Zhou et al. [187] proposed a Bayesian framework based on a simple human intuition which assumes that all humans have a head and two legs and each leg is joined at the knee. A 2D articulated model which is a crude approximation to a real walker is fitted to gait silhouettes. The gait images were manually labeled to find out sections of gait cycles. The objective of this approach is to determine the likelihood of the image given the model. The authors claim that their approach tackles well uncertainties such as occlusion and noise.

In most cases, it appears that the silhouette is at least as efficient as the low-dimensional features that can be extracted from a silhouette. This is intuitively expected since the feature extraction step is a lossy operation, i.e., in general, the silhouette cannot be reconstructed from the feature. However, feature extraction could dramatically reduce the high complexity imposed by silhouette features. Hence in the pursuit of the majority of the Video sensor-based GR approaches, the silhouette feature provides a useful target for gait performance. Motivated by these approaches, we will propose silhouette-based GR approaches in Chapters 5 and 6 of this thesis.

4.3.2 Floor Sensor(FS)-based GR approaches

Basically in this category of approaches, a set of sensors are installed in the floor, and gait-related data are measured. The use of floor sensors for gait analysis is commonly employed by physiologists. Middleto et al. [201] presented a prototype system for acquisition of footfall data. Eventually the system has been designed to study GR by applying an alternate modality. The three main components of the system comprised: a large sensor mat, a hardware interface, and an analysis software. The system consisted of 1536 individual sensors, arranged in a $3m \times 0.5m$ rectangular strip with an individual sensor area of $3cm^2$. The sensor floor has been designed to operate at a sample rate of 22 Hz. The sensor itself uses a simple design inspired by computer keyboards and is made from low cost, off the shelf materials. Application of the sensor floor to a small database of 15 individuals has been performed. Three features have been extracted viz., stride length, stride cadence, and time on toe to time on heel ratio. Two of these measures have been adapted from video based gait recognition while the third is new to this analysis. These features proved sufficient to achieve an 80% recognition rate. In 60% of the subjects the heel to toe measure alone has been found to be sufficient to recognise their identity. Hence the authors argue that the dynamic behaviour of the foot is a potent biometric in itself. But further research with large number of subjects is required to justify this claim. The authors suggest that future efforts can concentrate on how to increase the sensor resolution whilst keeping the cost down and the construction process simple. Ground reaction forces generated during normal walking have recently been used to identify and/or classify individuals based upon the pattern of the forces observed over time [202]. Body mass extracted from vertical ground reaction forces has been recently experimented by Jenkins and Ellis [202]. A recognition rate of about 40% with a data set of less than 70 subjects is reported which shows that more research is required to make FS-based GR approaches appealing. Further, FS-based approaches rely on weakly identifying biometrics such as the actual body mass. These quantities can gradually change over a short period of time. Let us assume that such weak biometrics, fused with other reliable biometrics, were to be applied in biometric featured passports. This will lead to the process of periodically updating biometric featured passports, in intervals of months, which is not practicably feasible as per current biometric standards.

4.3.3 Wearable Sensor(WS)-based GR approaches

This category of GR approaches are relatively recent compared to the other two approaches discussed above [168]. Usually motion recording sensors are worn or attached to various places on the body of a subject such as shoe and waist [203–205]. Examples of the recording sensors can be accelerometer, gyro sensors, force sensors, bend sensors, and so on that can measure various characteristics of walking. The movement signal recorded by such sensors is then utilized for person recognition purposes. Previously, the WS-based gait analysis has been used successfully in clinical and medical settings to study and monitor patients with different locomotion disorders [206]. In medical settings, such approaches are considered to be cheap and portable, compared to the stationary vision based systems [207]. Despite successful application of WS-based gait analysis in clinical settings, only recently these approaches have been applied for person recognition [204]. Unlike Video Sensor-based GR, no public data-set on WS-based gait is available, which makes it difficult for researchers to do empirical evaluations on a common dataset.

4.4 Summary

An insight into gait analysis techniques from diverse fields such as medical, biomechanics and psychology will aid to enhance machine vision based GR techniques. Decades back, the results of psychological studies revealed that people can recognize their friends from motion but motion alone is not sufficient to be a reliable form of identification. Lessons learnt from biomechanical literature infer that various gait components contribute to the overall gait recognition process in various proportions. However, various components of the gait of a subject must share a common frequency. There are motions at different frequencies within a gait. However, the overall gait has a fundamental frequency that corresponds to the complete gait cycle. Other frequencies are multiples of the fundamental frequency. This phenomenon is termed as frequency entrainment which infers that it is not possible to walk with component motions at arbitrary frequencies. Recent advances in computing technology, especially high-speed processors, bulk storage and memory resources, are enabling gait as a practicable biometric right now, although the idea has emerged decades ago. In the class of object recognition problems, identification is considered to be harder than verification.

Static parameter-based based approaches have not reported good performance on common databases, partly due to their need for 3D calibration information. Recognizing gaits under outdoor settings is more difficult than that of indoor settings.

Several studies conclude that shape is more significant than dynamics (kinematics) for person identification as dynamics is vulnerable to uncertainty caused as a result of high intrasubject variability. Silhouette-based gait recognition techniques are gaining much interest among current gait recognition researchers. The reason is that they do not need any further information such as color, texture or gray-scale metrics and yet they capture the motion of most of the body parts. Recent studies have shown that silhouette shape has equal, if not more, recognition potential than gait kinematics. This has inspired us to propose silhouette-based GR approaches, which will be presented in Chapters 5 and 6.

Chapter 5

Extending PROCIM for Robust Gait Recognition

5.1 Introduction

In this chapter we intend to propose a probabilistic sub-gait interpretation model to recognize gaits. A sub-gait is defined by us as part of the silhouette of a moving body. Binary silhouettes of gait video sequences form the basic input of our approach. A novel modular training scheme has been introduced in this chapter to efficiently learn subtle sub-gait characteristics from the gait domain. For a given gait sequence, we get useful information from the sub-gaits by identifying and exploiting intrinsic relationships using Bayesian networks. Finally, by incorporating efficient inference strategies, robust decisions are made for recognizing gaits. Our results show that the proposed model tackles well the uncertainties imposed by typical covariate factors and shows significant recognition performance.

Our **PRO**babilistic **C**omponent **I**nterpretation **M**odel which we abbreviate as (**PROCIM**) is based on a fundamental insight about human pattern matching and memory. While reasoning with objects which are prone to uncertainties, in our case visual processing of gaits, humans are often able to notice similarities between sub-gaits and gaits. When we see a person at a distance, we may notice a particular pattern of arm-swinging or hip movement as a characteristic of the whole walking gait of that person. A formal definition of sub-gait is given in equations (2.6)-(3.11) of Section 5.2. First a set of sparse components or sub-gaits of the cluttered gait pattern is perceived, this is the *probe*. These are then matched to a bulk set of gait patterns, the *gallery*, that are remembered. This reasoning based on similarity mapping is processed in such a way to reveal inherent conditional independencies between gaits. In our study we intend to scientifically represent these independencies using Bayesian Networks (BN). BNs serve as fundamental tools in tackling uncertainty problems as they characterize intuitive notions of human reasoning. In other words, PROCIM employs BNs to find out and learn intrinsic sub-gait mappings that naturally exist in gait patterns. We derive robust probabilistic decisions by exploiting the mappings established. We have identified three potential sub-gaits among the possible sub-gaits of a gait silhouette, by experimental evaluation. Selecting potential sub-gaits is based on how significantly they contribute to the recognition mechanism of gaits. We will provide details in Section 5.5.2.

We briefly present PROCIM's architecture with the aid of the flow diagram shown in Fig.5.1. Firstly we decompose the gallery silhouettes into sub-gaits and subject them to an appropriate feature extraction process to construct a low dimensional feature space. PROCIM is a generic model which could be applied to any feature space (subspace) projection technique, such as PCA or SVM. For demonstration sake we have used the recently proposed MPCA feature space [196].

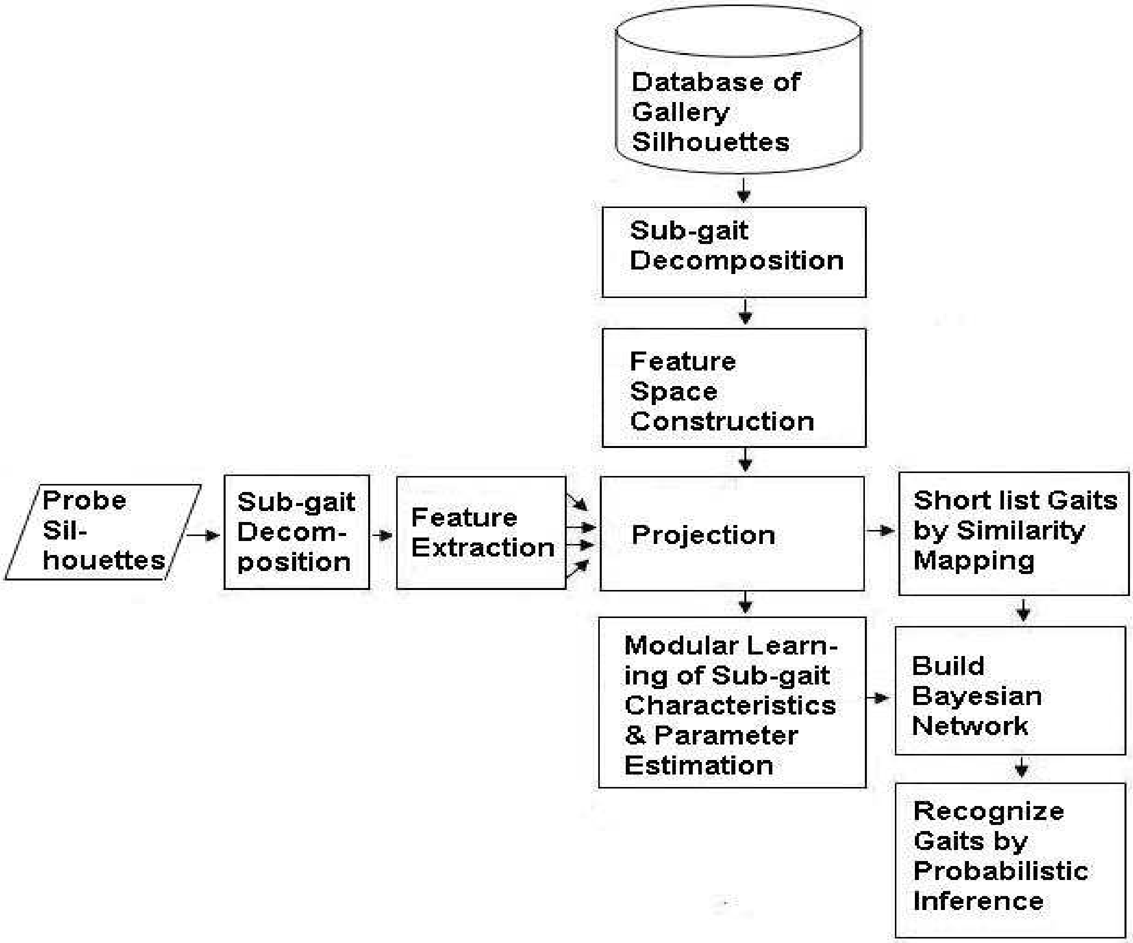

Database of Gallery Silhouettes
Sub-gait Decomposition
Feature Space Construction
Probe Sil-houettes
Sub-gait Decom-position
Feature Extraction
Projection
Short list Gaits by Similarity Mapping
Modular Learn-ing of Sub-gait Characteristics & Parameter Estimation
Build Bayesian Network
Recognize Gaits by Probabilistic Inference

FIGURE 5.1: Framework of the proposed **PRO**babilistic **S**ub-gait **I**nterpretation **M**odel (**PROCIM**)

PROCIM further learns subtle sub-gait characteristics using a novel modular training scheme introduced in this study. Also using standard machine learning procedures, PROCIM estimates the parameters of the BN. All these preliminary activities are performed off-line to make minimal use of computing resources. Secondly the probe silhouettes are decomposed into similar sub-gaits and their extracted features are projected onto the feature space. Then gaits are shortlisted with the aid of similarity mapping-based reasoning. The intrinsic relationships between the sub-gaits and gaits are represented intuitively using BNs. Finally gaits are recognized by exploiting these relationships using robust probabilistic inference techniques.

The rest of the chapter is organized as follows. Section 5.2 formally present the proposed sub-gait segmentation scheme. We propose a novel modular training scheme in Section 5.3. Section 5.4 demonstrates the robustness of PROCIM against some common variations. We discus experimental results and evaluate its performance in Section 5.5. Section 5.6 summarizes this chapter.

5.2 Sub-gait Segmentation

Segmenting specific body components such as head, torso, arms and legs demands manual labeling. However, manual labeling may not guarantee accurate marking of the body components on video sequences. This is because of factors such as low-image quality due to overall intensity, occlusion of feet when walking on grass, similarity of dark skin tones of some subjects with the background, occlusion of the arms due to various viewing angles, and the presence of dark or baggy clothing[164]. We intend to avoid such manual labeling and at the same time utilize the information from those body components. Hence we strategically segment the silhouettes into sub-gaits viz., Upper Gait(U), Mid Gait (M), Lower Gait(L), LeFt Gait(LF) and Right Gait (R). We will represent the set of sub-gaits by $S = \{U, M, L, LF, R\}$. By manipulating the binary files that represent silhouettes, we compute the bounding rectangle that encompasses a silhouette and resize them to a standard dimension of 64 x 44 pixels. A typical silhouette frame of a gait video sequence and its sub-gaits are shown in Fig.5.2.

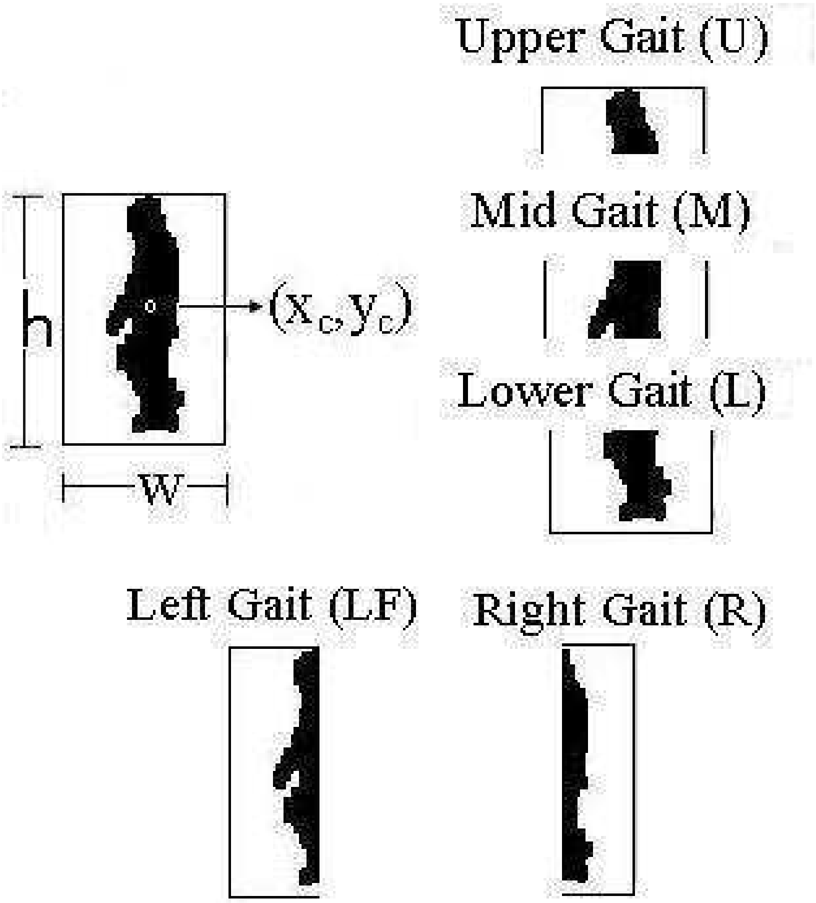


FIGURE 5.2: A typical silhouette and its sub-gaits. A sub-gait is defined by us as part of the silhouette of a moving body. Specifically all the sub-gaits viz., U, M, L, LF and R, are defined in equations 5.2 to 5.6.

We define these sub-gaits using the language of set theory which is widely used to represent and describe image semantics [146]. For a given silhouette frame I(x,y) with width w and height h, its centre (x_c, y_c) can be computed by

$$(x_c, y_c) = \left(\frac{w}{2}, \frac{h}{2}\right). \tag{5.1}$$

Then the sub-gaits U, M and L can be defined as

$U(I(x,y)) = \{(x,y)|x_c - \frac{w}{2} \leq x \leq x_c + \frac{w}{2},$

$$y_c + \frac{h}{2} \leq y \leq y_c + \frac{h}{2} - h\epsilon_1\}, \tag{5.2}$$

$$M(I(x,y)) = \{(x,y)|x_c - \tfrac{w}{2} \leq x \leq x_c + \tfrac{w}{2},$$

$$y_c + \frac{h}{2} - h\epsilon_1 < y \leq y_c + \frac{h}{2} - h(\epsilon_1 + \epsilon_2)\}, \tag{5.3}$$

$$L(I(x,y)) = \{(x,y)|x_c - \tfrac{w}{2} \leq x \leq x_c + \tfrac{w}{2},$$

$$y_c + \frac{h}{2} - h(\epsilon_1 + \epsilon_2) < y \leq y_c - \frac{h}{2}\}. \tag{5.4}$$

For the sub-gait definitions above, the heights of each of the sub-gait segments are distinct and determined by constants ϵ_1 and ϵ_2. Values of these constants were chosen based on rough estimates performed on the mean silhouette of the gallery set. The left and right sub-gaits viz., LF and R, which are segmented from the centre are just a function of width (w) and do not require extra constants. Hence their definitions are straight forward as follows:

$$LF(I(x,y)) = \{(x,y)|x_c - \tfrac{w}{2} \leq x \leq x_c,$$

$$y_c + \frac{h}{2} \leq y \leq y_c - \frac{h}{2}\}, \tag{5.5}$$

$$R(I(x,y)) = \{(x,y)|x_c < x \leq x_c + \tfrac{w}{2},$$

$$y_c + \frac{h}{2} \leq y \leq y_c - \frac{h}{2}\}. \tag{5.6}$$

In this chapter we describe a procedure to recognize gaits by representing and interpreting sub-gait characteristics using a reasonable probabilistic framework. Finding optimal sub-gait dimensions such as the optimal height of the sub-gaits might further improve recognition performance. Such operational motivation factors needs further scrutiny of rigorous iterative experiments, exploration of advanced image segmentation and optimization techniques which deserve another dedicated study. We employ BNs to establish intrinsic similarity mappings between the sub-gaits and the gaits. Each node of a BN has a set of probable values for each variable which are known as belief states. These belief states are propagated between nodes of the BN effectively. A BN maps intrinsic relationships that are inherent in a domain in terms of parent and child nodes. It is capable of learning these relationships and storing the belief states of a given domain in the form of Conditional Probability Tables (CPT). By manipulating these belief states, the state of a particular node can be queried from other nodes with the aid of probabilistic inference techniques. In our case we would like to query the belief state of a gait sequence by observing the probabilities entailed by its sub-gaits.

Though gait motion is periodic in nature, various sub-gaits contain different information about the gait they constitute. Owing to this variation, all the sub-gaits will not have the same probability of influencing the gait to be recognized. Therefore each sub-gait of a gait will have different belief states and this varies from subject to subject as the walking style of individuals varies. The more unique features a sub-gait contains, the more strength it will have to influence the recognition of the gait to which the sub-gait belongs. We define the strength of a sub-gait S_i which crucially contributes to the recognition of the gait G_p as *Influence Strength* and denote it as Z_{ip}. We will call the parent nodes of the proposed BN the prior belief states of the sub-gaits. The gaits influenced by the sub-gaits are the child nodes of the BN. In this probabilistic framework we will infer the belief state of the gaits conditional on the sub-gaits, in order to recognize the gaits.

5.3 Proposed Modular Training Scheme

We will describe a novel modular training scheme employed by PROCIM here. A training or test sample is well defined in many object recognition (eg. face, iris recognition) problems . For example, a face or an iris image is considered as a sample without any further partitions. However, the definition of a gait sample is subjective and not so precisely defined. Usually a gait sample is represented in terms of gait cycles (either full, multiple or partial cycles). A gait cycle begins when one foot contacts the ground and ends when the same foot contacts the ground again. Thus, each cycle begins at initial contact with a stance phase and proceeds through a swing phase until the cycle ends with the next initial contact of the limb. Prior to factoring the gait samples into modules, we have constructed the sub-gaits data sets from the gallery data sets by applying the sub-gait segmentation scheme formulated in Section 5.2. In the gallery set, because each subject's behavior is represented as several gait samples due to variations in walking speed, the number of frames per sample will be different. A suitable time mode normalization algorithm can be applied to normalize the gait samples to have a unique number of frames. We have normalized the number of frames in each sample by applying the time mode normalization technique proposed by [196]. We intend to decompose the normalized sub-gait samples into compact modules and train PROCIM to learn the intrinsic relationships between these modularized sub-gaits. The proposed modular training scheme enables PROCIM to represent and learn subtle walking patterns of human gaits.

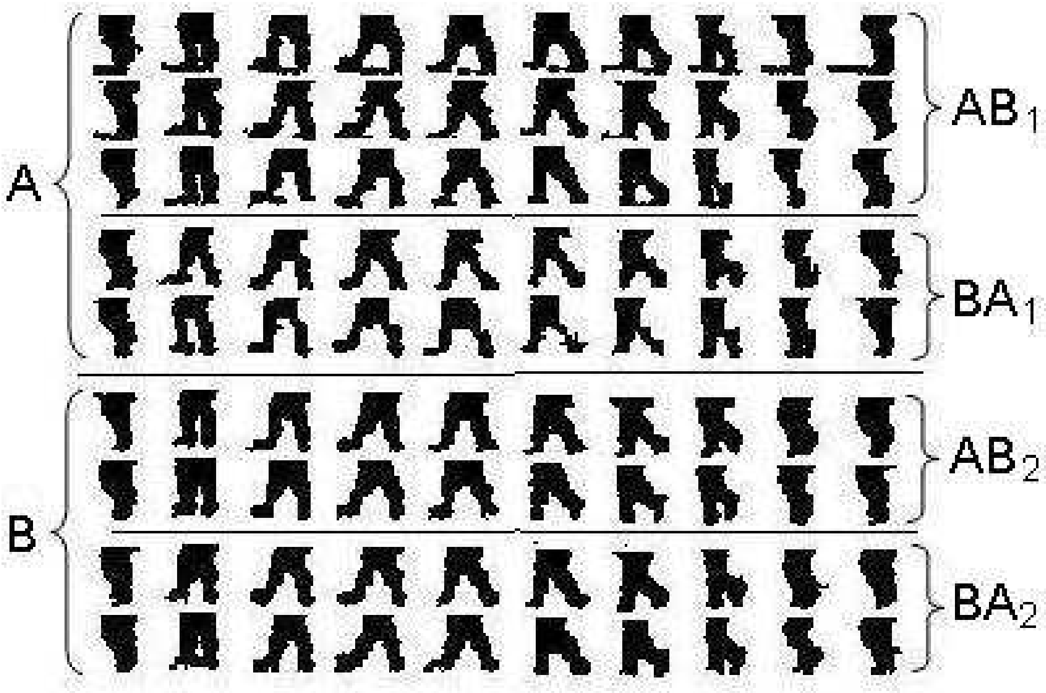


Figure 5.3: Modular scheme of a typical sub-gait

Fig.5.3 shows the modular scheme applied to a typical sub-gait. For example's sake we have shown the scheme for a lower sub-gait. We initially modularize all training samples into two subsets namely A and B. An even number of samples is split 50-50, an odd number the closest integer partition to 50-50. Gait subsamples of modules A and B represent how the subjects walk during the first part and second part of a walking segment. We further modularize these subsets into AB and BA which will have mixtures of walking samples from A and B together. That is AB will have some samples from the first half of A and B and BA will have some samples from the second half of A and B. Finally we modularize AB and BA into tiny modules viz., AB_1, AB_2, BA_1 and BA_2. That is each of these tiny modules represent about a quarter of a sub-gait sample. Mathematically we can model this modular scheme as follows:

Let the gallery set of sub-gait (or gait) silhouettes of a subject say U be represented by d gait samples. Let the i^{th} sub-gait sample be denoted by u_i, where $1 \leq i \leq d$. We wish to modularize U such that

$$U = AB_1 \cup BA_1 \cup AB_2 \cup BA_2 \tag{5.7}$$

where

$$AB_1 = \sum_{i=1}^{a} u_i; \quad BA_1 = \sum_{i=a+1}^{b} u_i \tag{5.8}$$

$$AB_2 = \sum_{i=b+1}^{c} u_i; \quad BA_2 = \sum_{i=c+1}^{d} u_i \tag{5.9}$$

The indices a, b and c of Eqs. (3.24) and (5.9) can be computed as

$$a = \lceil d/4 \rceil; \quad b = a + \frac{d-a}{3}; \quad c = b + \frac{d-a}{3} \tag{5.10}$$

Obviously the tiny modules defined in (3.24) and (5.9) can be appropriately merged to yield

$$AB = AB_1 \cup AB_2; \quad BA = BA_1 \cup BA_2 \tag{5.11}$$

Modules can be combined using the following rules

$$AB \cap A = AB_1; \quad BA \cap A = BA_1 \tag{5.12}$$

$$AB \cap B = AB_2; \quad BA \cap B = BA_2 \tag{5.13}$$

We will show shortly how the proposed modular scheme enables us to relate the various sub-gaits and learn subtle walking patterns that are inherent in a subject's walking behavior. We perceive that the intrinsic relationships that exist between the modularized sub-gaits contribute significantly in governing the gait patterns. The BN employed in PROCIM learns the belief states of these relationships systematically from the sub-gait data sets using the MLE approach outlined in Section 3.3.1. The learned belief states are stored in the form of Conditional Probability Tables (CPTs). For k sub-gaits and m modules, the BN yields a CPT comprising of $2^{k*m}-1$ number of rows. A typical CPT for the case of two sub-gaits L and LF whose samples are factored into two subsamples A and B is shown in Table 5.1.

TABLE 5.1: CPT showing belief states of subtle sub-gait relationships learned from the proposed modular training scheme for some typical subjects. The sub-gait operators $L(\cdot)$ and $LF(\cdot)$ have been defined in equations (5.4) and (5.5).

Sub-gaits	Learned belief states of typical subjects				
	G_5	G_{10}	G_{15}	G_{20}	G_{25}
L(A)	0.97	0.86	0.75	0.93	0.86
L(B)	0.79	0.84	0.73	0.79	0.84
L(A) L(B)	0.78	0.99	0.92	0.99	0.99
LF(A)	0.94	0.86	0.89	0.95	0.86
L(A) LF(A)	0.97	0.86	0.68	0.79	0.86
L(B) LF(A)	0.78	0.97	0.67	0.99	0.97
L(A) L(B) LF(A)	0.93	0.58	0.69	0.91	0.58
LF(B)	0.92	0.87	0.86	0.97	0.87
L(A) LF(B)	0.97	0.78	0.83	0.99	0.78
L(B) LF(B)	0.97	0.66	0.85	0.97	0.66
L(A) L(B) LF(B)	0.56	0.58	0.33	0.77	0.58
LF(A) LF(B)	0.75	0.99	0.92	0.99	0.99
L(A) LF(A) LF(B)	0.93	0.58	0.69	0.91	0.58
L(B) LF(A) LF(B)	0.58	0.73	0.39	0.72	0.73
L(A) L(B) LF(A) LF(B)	0.97	0.78	0.83	0.99	0.78

By combining various sub-gait modules we can reveal intrinsic characteristics of gait patterns. For example the CPT entry $L(A)\,LF(B)$ intends to reveal the belief state of "*left leg sub-gait pattern*" for a portion of a walking sequence. Trivially $L(A) \cap LF(B) = L(LF(AB))$. When more combinations of sub-gaits and sub-samples are involved, the interpretation needs a few more steps. For example a typical CPT entry and its interpretation are as follows:

$L(A) \quad L(B) \quad LF(A)$

$= L(A) \cap L(B) \cap LF(A)$

$= L(AB) \cap LF(A)$

$= L(LF(AB_1)) \because Eq.(5.12)$

Similar logical reasoning can be extended to interpret any other entry in the CPT. The conditional probabilities in Table 5.1 give a measure of the strength of sub-gait relationships. For example, referring to the first column in the table, we observe that the conditional probabilities $P(G_5|L(A))$, $P(G_5|L(A), LF(A))$ and $P(G_5|L(A), LF(B))$ are higher. This reveals the fact that the gait motion of the subject G_5 is highly characterized by these intrinsic sub-gait relationships. Whereas $P(G_{15}|L(A), L(B), LF(B))$ and $P(G_{15}|L(B), LF(A),$ $LF(B))$ (middle column of the table), being low indicate that G_{15} is poorly characterized by these sub-gait modules. We will shortly see how robust probabilistic decisions can be made by interpreting and exploiting these subtle relationships.

5.4 Robustness to Common Variations

Some common uncertainties encountered in the process of gait recognition are caused due to variations present in challenging outdoor environments such as view, surface, shoe, missing body components and so on. The experimental results of PROCIM's robustness against these uncertainties will be presented in Section 4. Here we will analyse the effect of uncertainties caused by two typical variations viz. view and missing body components. The scenario of a typical probe gait whose gallery representation is $G58$ has been subjected to viewing variations of 18° and 162° are shown in Fig. 5.4 and Fig. 5.5 respectively. The Bayesian Network (BN) generated by PROCIM (Fig. 5.4 and Fig. 5.5) helps us to analyse how this uncertainty affect the recognition mechanism, in particular the relationships between the gaits and sub-gaits. As the viewing variation of 18° is relatively small and probably other variations being less severe, all the sub-gaits of $G58$ collectively contribute to the recognition process as seen in Fig. 5.4. Further the silhouettes are noisy due to factors such as similarity of colors of the subject and the background, varying illumination caused by the operating environment and so on. Despite these variations, $G58$ has been successfully recognized as a winner gait as shown in the bar chart.

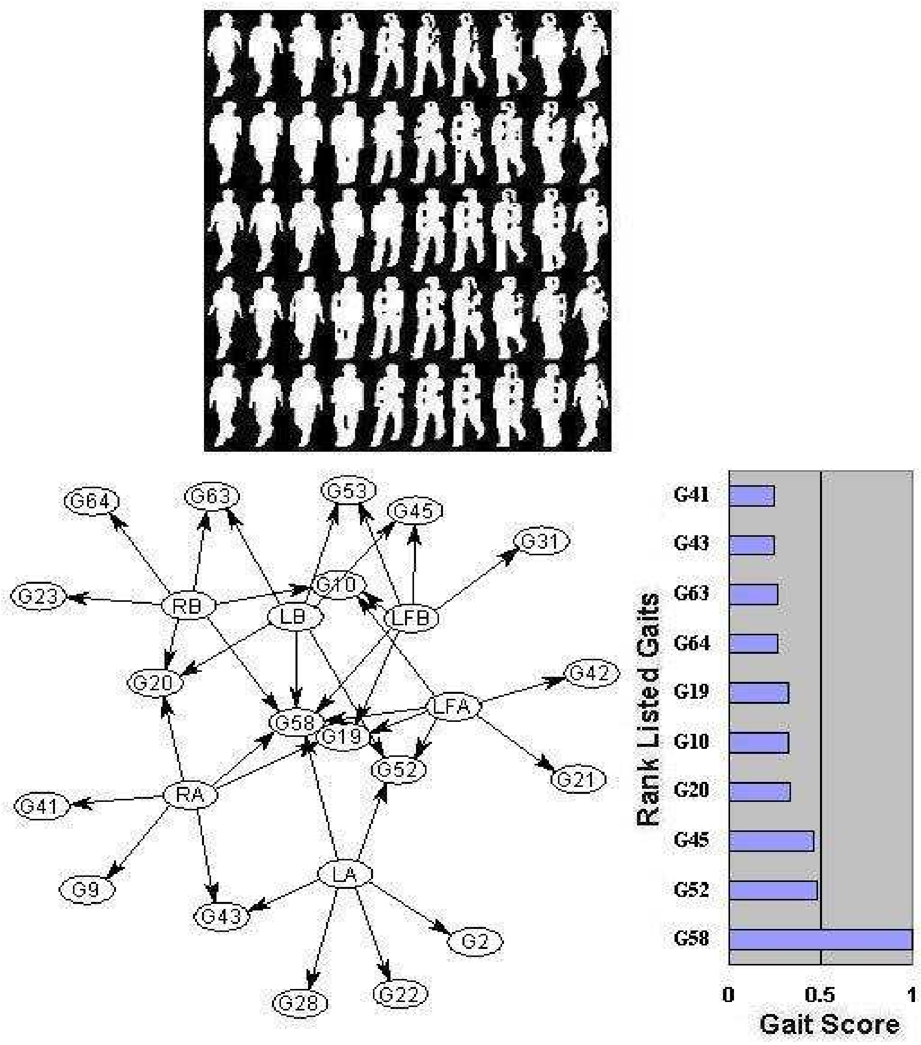


FIGURE 5.4: A scenario that depicts the recognition process of a probe gait (typically chosen from CASIA dataset) with a typical viewing variation of 18°. Normalized silhouettes of gait sequences of the probe (each row represent one sample), the associated Bayesian Network generated by PROCIM and the bar chart of first ten winner gaits being recognized are shown.

Body components such as head, arms and some portion of the torso are missing in most of the normalized silhouette sequences shown in Fig. 5.5.A huge viewing variation of 162° along with the complexity of missing body components, obviously causes more uncertainty and consequently $G58$ has been degraded from rank1 to rank2 as shown in the the bar chart.

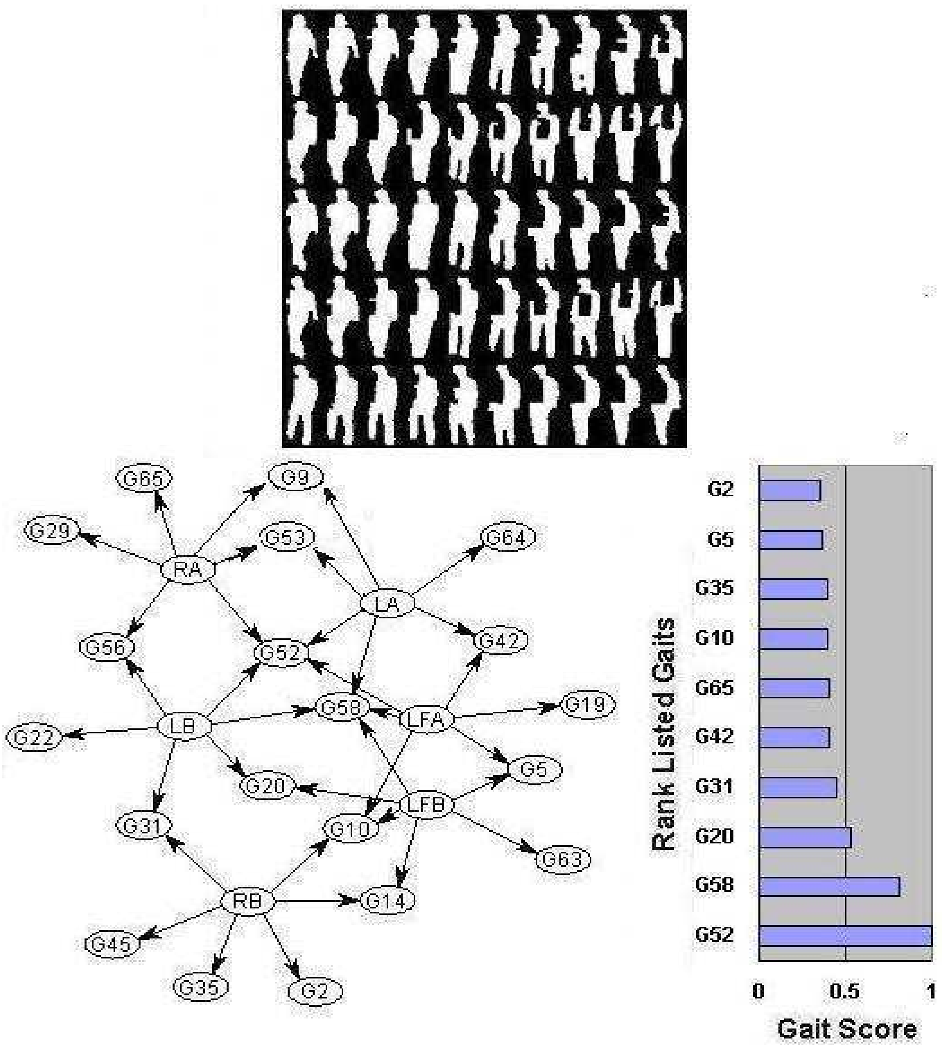


Figure 5.5: A scenario that depicts the recognition process of a probe gait (typically chosen from CASIA dataset) with a typical viewing variation of 162°. Normalized silhouettes of gait sequences of the probe with missing body components, the associated Bayesian Network generated by PROCIM and the bar chart of first ten winner gaits being recognized are shown.

Interestingly when the gait of a subject is viewed from 162°, the left body motion is more visible than the right body motion. This is intuitively reflected by the sub-gait to gait relationships captured by the BN shown in Fig. 5.5. Specifically the right sub-gaits, RA and RB, have not contributed to the recognition of $G58$. However these sub-gaits played their role when the viewing angle was 18° as seen in Fig. 5.4. The proposed framework enables us to visualize such interesting relationships that exists between gaits and sub-gaits. We see that sub-gaits RA and RB lack to provide evidence due to uncertainties in the scenario. However PROCIM grasps information by accumulating evidences from other sub-gaits. By manipulating the available evidences (LA, LB, LFA and LFB) and the learned belief states from the stored CPTs, PROCIM is still able to recognize $G58$ reasonably well (in second rank).

The gait samples of a subject is represented in terms of normalized gait cycles which comprises a set of silhouettes. Some of the samples might have silhouettes with missing parts (weak samples). Within a sample the uncertainty caused by silhouettes with missing parts will be compensated by the ones which are complete. Furthermore, as we decompose the samples into compact modules (please see Section 5.3), the modules which have more good samples would compensate the uncertainty caused by modules that contain silhouettes with missing parts. For example a module of the left sub-gait (LFA) might fail to provide evidence or provide less evidence (influence strength could be weak due to weak samples). However the other module of the left sub-gait (LFB) or modules of other sub-gaits might provide sufficient evidence to mitigate the uncertainties imposed by the weak module.

5.5 Experimental Validation

5.5.1 Data set and experimental design

We have used the University of South Florida (USF) HumanID gait challenge data set [12] and the multi-view gait dataset offered by Chinese Academy of Sciences [194] to evaluate PROCIM and compared it with the state-of-the-art gait recognition algorithms. The USF data set which was collected on typical outdoor environment, consists of 122 subjects comprising 1870 video sequences. The gait challenge baseline algorithm [12] as well as very recent algorithms such as [173] consider seven standard experimental probe sets, the details of which are tabulated in Table 5.2.

Table 5.2: Experimental Notation and Description with compliance to Human Identification in USF HumanID Data Sets

Probe Set	Capturing Condition	Covariate Factors
A	GAL	View
B	GBR	Shoe
C	GBL	Shoe, View
D	CAR	Surface
E	CBR	Surface, Shoe
F	CAL	Surface, View
G	CBL	Surface, Shoe, View

The seven probe sets, A to G, are designed to perform a range of experiments in the order of increasing difficulties. The abbreviations of various capturing conditions in the table viz., $C, G, A, B, L,$ and R refers to Concrete surface, Grass surface, shoe type A, shoe type B, Left view and Right view respectively.

5.5.2 Identifying Potential Sub-gaits

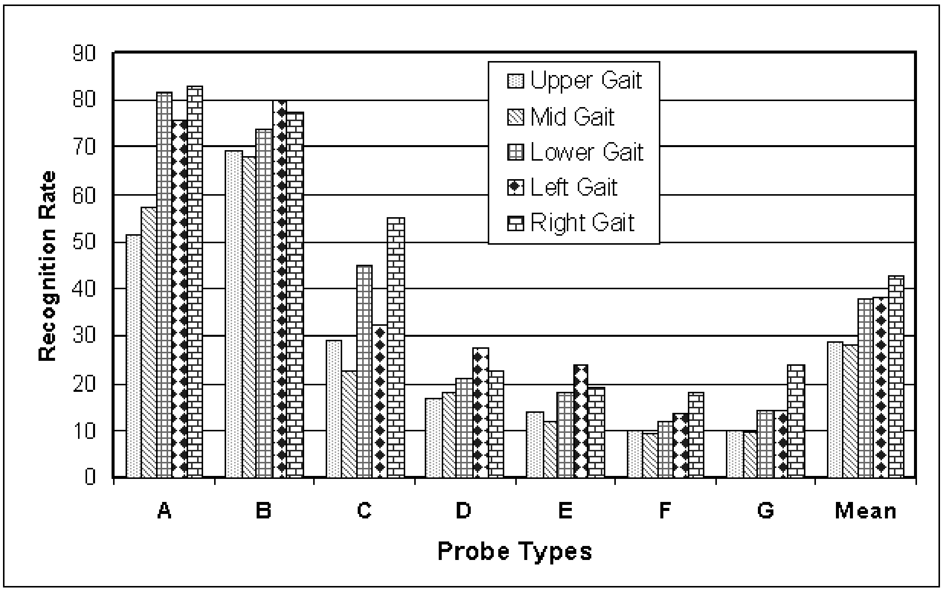


Figure 5.6: Recognition potential of sub-gaits for the HumanID gait challenge data

Recall from Section 5.2 that we have defined five sub-gaits ($k = 5$) viz., Upper Gait(U), Mid Gait(M), Lower Gait (L), Left Gait(LF) and Right Gait (R). Also recall from Section 3.3.2 that the size of the Bayesian Network tends to grow exponentially as the number of sub-gaits increases. This will in turn demand more computing resources. Hence prior to parameter estimation, strategically selecting just a few potential sub-gaits would enable PROCIM to be computationally feasible. In this section, we will identify such potential sub-gaits based on their recognition power. We have computed recognition rates for all of the sub-gaits for the seven core experiments using the approach proposed by [196], the results of which are shown in Fig.5.6. The mean performance of all the experiments shown at the right most end of Fig.5.6 justifies that the sub-gaits L, LF and R have higher recognition potential than U and M. Hence we will only employ these potential sub-gaits in the subsequent experiments.

5.5.3 Comparison of PROCIM with state-of-the-art

We have experimented PROCIM with the HumanID gait challenge experiments by gradually increasing the number of sub-gaits. A Cumulative Match Characteristic (CMC) curve [208] shows various probabilities of recognizing an individual depending on how similar their measurements are to that of others in the gallery. The rank 1 point on the CMC curve is the nearest-neighbor recognition performance. The CMC graphs of these experiments are shown in Fig.5.7,5.8,5.9 and 5.10. Initially by considering the lower sub-gait alone (i.e. L), mean recognition rates of about 52% and 62% have been yielded by PROCIM respectively for the rank1 and rank5 performance. Then by combining two potential sub-gaits (i.e $L + LF$), this improved to about 60% and 76%. Finally by considering all the three potential sub-gaits (i.e. $L + LF + R$), the mean recognition rates have been considerably improved to about 69% and 85% for rank1 and rank5 performance respectively. These experimental results clearly show that when all the potential sub-gaits are used, PROCIM achieves maximum recognition performance.

Further we subject PROCIM to the HumanID gait challenge experiments using

USF dataset and compared it against the following state-of-the-art gait recognition algorithms:

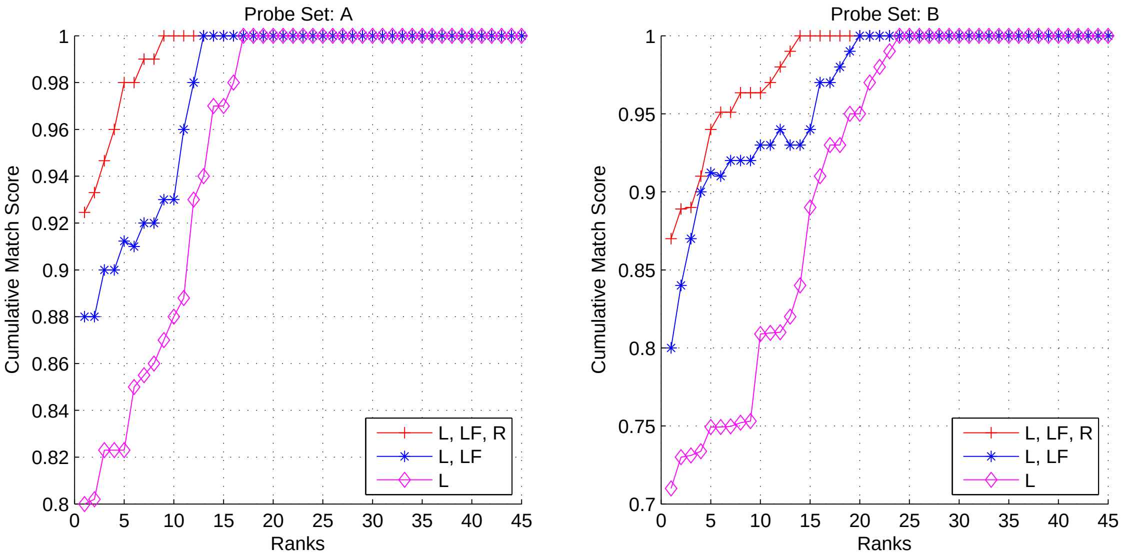


FIGURE 5.7: CMC response of PROCIM with respect to HumanID gait challenge experiments A and B

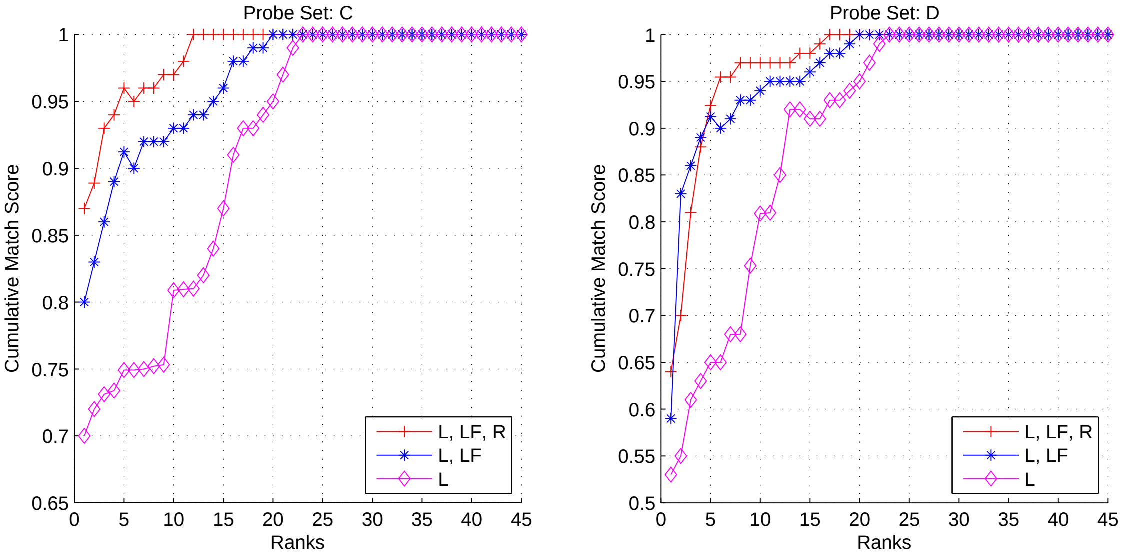


FIGURE 5.8: CMC response of PROCIM with respect to HumanID gait challenge experiments C and D

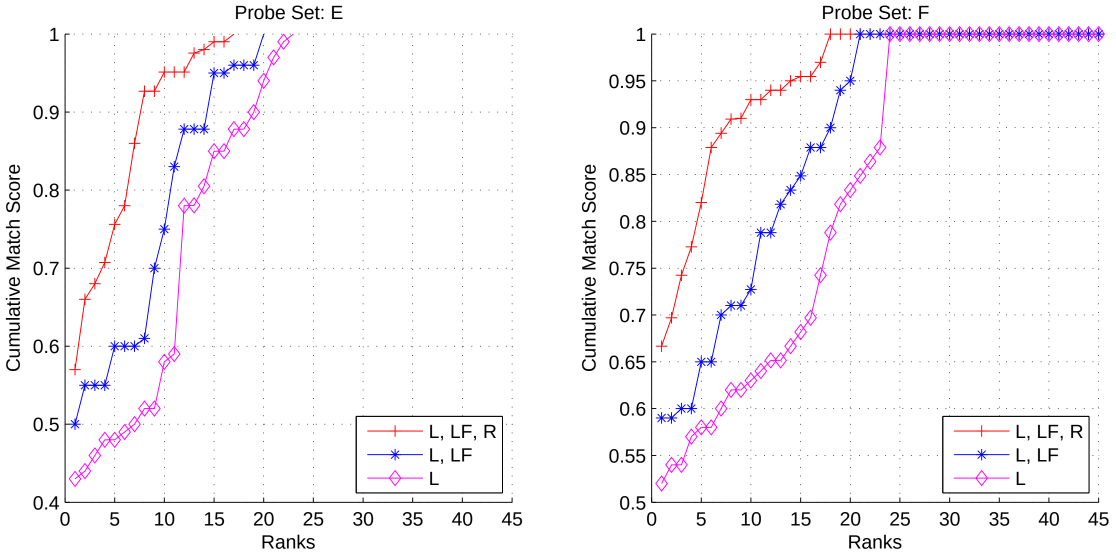


FIGURE 5.9: CMC response of PROCIM with respect to HumanID gait challenge experiments E and F

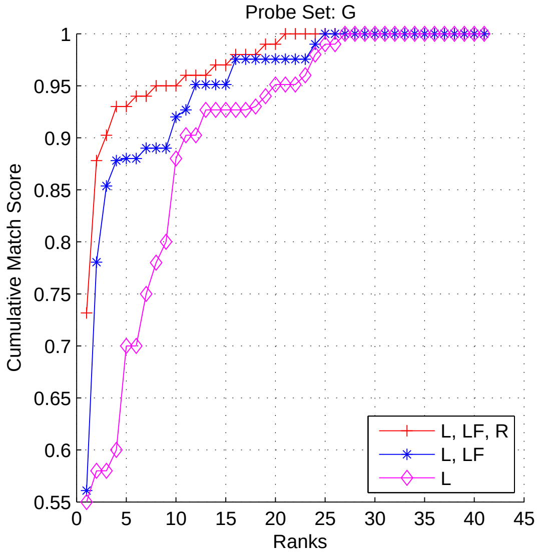


FIGURE 5.10: CMC response of PROCIM with respect to HumanID gait challenge experiments G

i. Baseline [12]

ii. HMM - Hidden Markov Model [188]

iii. DATER - Discriminant Analysis with TEnsor Representation [209]

iv. DTW/HMM - Dynamic Time Warpring/HMM [198]

v. ETGLDA - Eigen Tensor Gaits based on Linear Discriminant Analysis [196]

vi. GEI - Gait Energy Image [210]

vii. LTN - Linear Time Normalization [195]

viii. MR - Matrix Representation [211]

ix. NTWN - Nonlinear Time-Warp Normalization [173]

x pHMM - population Hidden Markov Model [9]

We have experimented PROCIM with two modes of recognition experiments. Initially we used the conventional experimental setting proposed by [12] where training was done with a limited gallery set (capturing condition was fixed as Grass, Shoe Type A and Right Camera). Recognition tests were performed with various probe sets which are described in Table 5.2. We refer to this conventional recognition experiment as PROCIM-a. Very recently [173] have shown that improved recognition rates can be achieved by using multiple samples for training. They proposed a round-robin recognition experiment in which one of the challenge sets was used as test while the other seven were used as training examples. The process was repeated for each of the seven challenge sets. We refer this experiment as PROCIM-b.

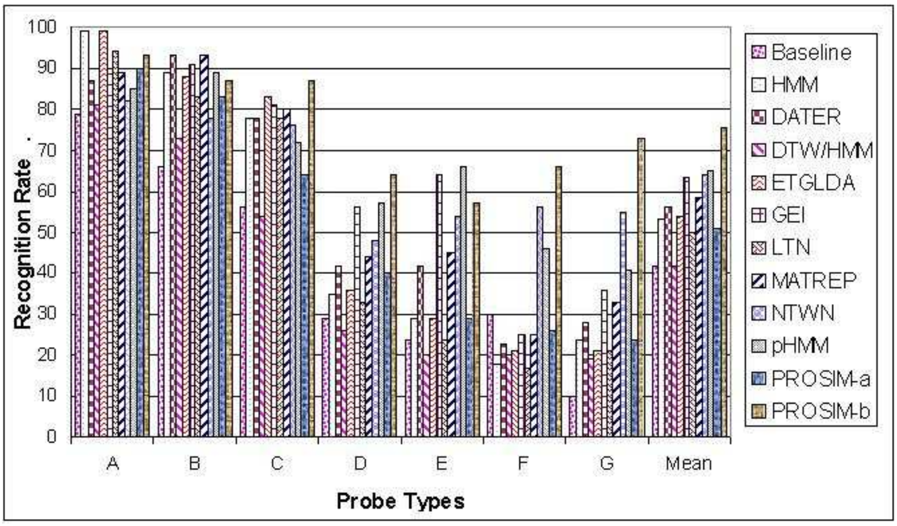


FIGURE 5.11: PROCIM Vs. state-of-the-art gait recognizers: Rank1 Performance

The rank1 and rank5 performance comparisons with state-art-of-the-art gait recognition algorithms are shown as bar charts in Fig. 5.11 and Fig. 5.12 respectively. Though PROCIM-a competes fairly with other algorithms, it is not as significant as PROCIM-b due to the restricted mode training. We see that PROCIM-b outperforms other algorithms in majority of the tests. Recognition rates of 75.3% and 89.6% achieved by PROCIM-b respectively for rank1 and rank5 performance, on an average of all the seven gait-challenge experiments, justifies the robustness of the proposed approach.

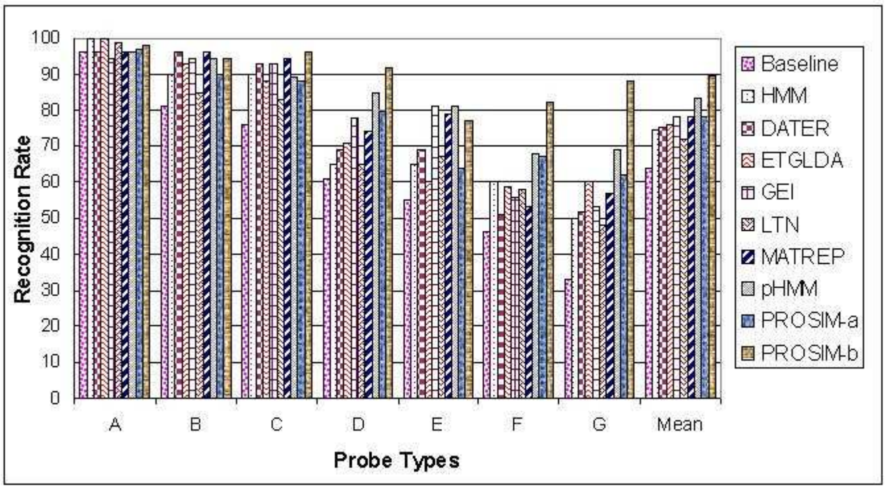


Figure 5.12: PROCIM Vs. state-of-the-art gait recognizers: Rank5 Performance

5.5.4 Experiments with the CASIA dataset

In this section we will investigate the generalization capability of PROCIM with the large multi-view CASIA dataset which contains gait sequences of 124 subjects captured from 11 viewing angles. There were totally 10 gait sequences for each subject (6 normal + 2 with a coat + 2 with a bag) for each of the 11 views. The dataset [194] enables us to experiment the effect of the following co-variate factors.

i. View (Camera angles were varied from 0° to 180° in increments of 18°)

ii. View and clothing (i. + Subjects walked by covering them with a long coat)

iii. View and carrying (i. + Subjects walked by carrying a bag)

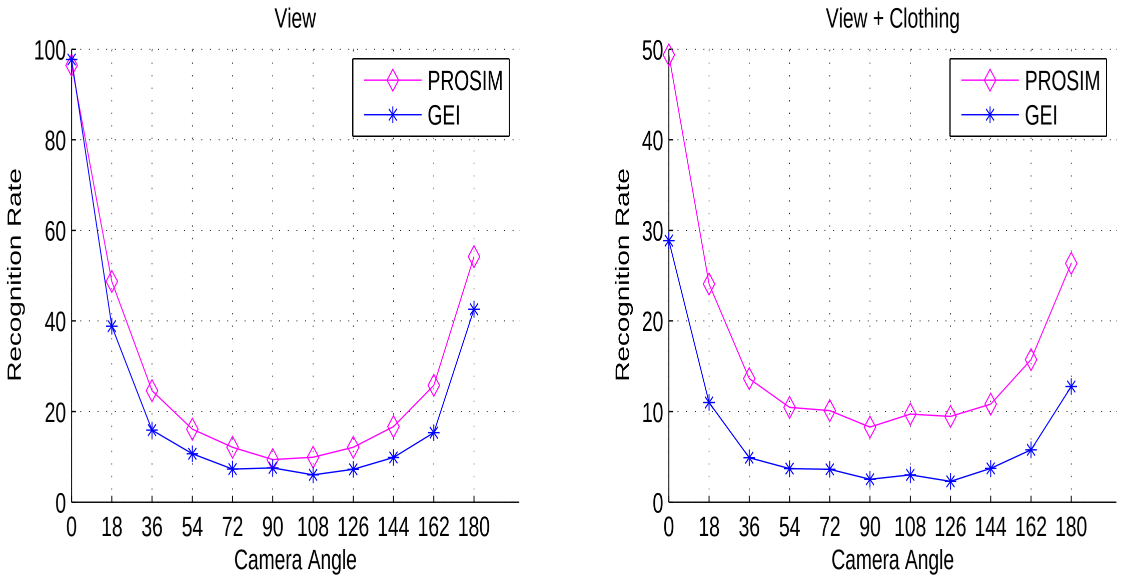


Figure 5.13: PROCIM and GEI algorithms are compared using the CASIA dataset; Cumulative match scores for typical variations, View and View+Clothing, are compared for the two algorithms.

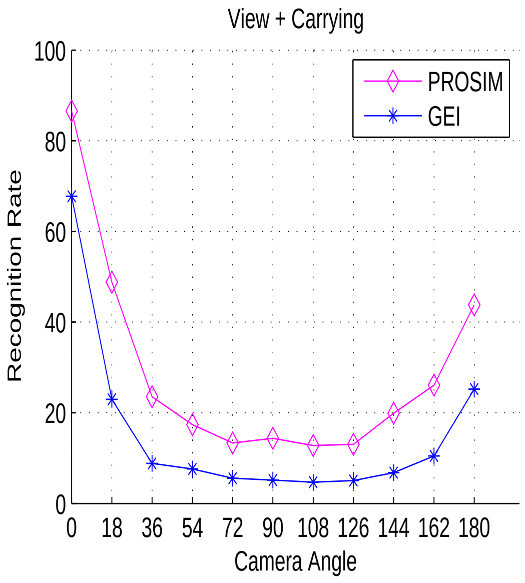


Figure 5.14: PROCIM and GEI algorithms are compared using the CASIA dataset; Cumulative match scores for typical variations, View+Carrying, are compared for the two algorithms.

The first four sequences (normal) were used for training and the remaining were used for testing. Yu et al. [212] have implemented the GEI algorithm using the CASIA dataset. We have compared our results with the GEI algorithm which are shown in Fig. 5.13 and Fig. 5.14. When tested by varying the carrying condition alone (i.e. for the same view), PROCIM and GEI yielded a recognition rate of about 87% and 68% respectively. When tested by varying the clothing condition alone, PROCIM and GEI yielded a recognition rate of about 50% and 29% respectively. This indicates that clothing is a tough test as the occlusion caused by long coat (most of the body parts are occluded by a long coat) imposes vast uncertainty to the recognition process. For a small viewing variation of 18°, PROCIM and GEI yielded a recognition rate of about 49% and 39% respectively. However when viewing is varied extremely (trained with 0° and tested with 90°) coupled with clothing variation the recognition rate has been degraded to 8.3% and 2.5% respectively by PROCIM and GEI. Significant improvement in performance has been achieved by PROCIM especially over cloting and carrying conditions.

5.6 Summary

We have identified potential sub-gaits and discovered interesting sub-gait characteristics within the gait domain. The novel Probabilistic Component Interpretation Model (PROCIM) introduced in this chapter does not require manual labeling of body components. Further the proposed modular training scheme enables PROCIM to learn subtle gait patterns. The graphical nature of PROCIM aids to intuitively visualize intrinsic sub-gait relationships and demonstrates how these sub-gaits collectively contribute to the recognition process. With the aid of few potential sub-gaits PROCIM reports a reliable recognition performance and competes well with the state-of-the-art gait recognizers. PROCIM is a generic model which can be fitted to suit any subspace technique. Our results show that extreme viewing angle variations coupled with change of clothing remains to be the toughest test among the experiments we have performed.

An interesting avenue for future directions could be "The proposed model does not have direct dependencies among parts and does this detract from the power of the modeling?". We have applied Bayesian Networks (which use directed edges) in the proposed framework to exploit the conditional independence properties that exists between gaits and their sub-gaits to achieve robust gait recognition. Such independence assumptions reduce the number of parameters in the model, and therefore making the model computationally feasible for real time applications. However setting dependencies among parts could be modeled using undirected links. Graphical models such as Markov networks [213] which use undirected graphs can be employed to capture dependency among various sub-gaits. In this regard, it will be an interesting avenue in the future to apply undirected graphical models, to investigate the impact of dependencies between sub-gaits and ultimately how they would influence the gait recognition process. Further we intend to apply the proposed approach to a wide range of object recognition problems in the future.

Chapter 6

Fusion based Gait Recognition: A basic Framework

6.1 Introduction

We introduced the PROCIM architecture in Chapter 3 to address the uncertainty issues imposed by the facial domain. Further in Chapter 5 we extended it to address the gait recognition problem which involves processing of dynamic images. PROCIM employs Bayesian Networks (BNs), which are basically directed acyclic graphs (DAGs). We have shown that the BN-based PROCIM architecture, exhibits several advantages including exploiting conditional independence properties exhibited by the domain and capable of making robust decisions under uncertainty. Such directed graphical models are useful because both the structure and the parameters provide a natural representation of real-world domains. However, the acyclic constraint of Bayesian networks does not permit it to express certain types of interactions. For example, let us consider a way to express an intuition such as "*Twins tend to have similar gaits*". This can be modeled in first-order logic using the following simple statement,

$$\forall x \forall y Twins(x, y) \implies (gait(x) \iff gait(y)),$$

which involves a bidirectional rule. Such bidirectional rules cannot be modeled using Bayesian Networks. Logic based paradigms are useful in modeling a variety of phenomena where one cannot naturally ascribe a directionality to the interaction between variables. First-order logic commits to the existence of objects and their relationships and enables means to express facts about *some* or *all* of the objects in a domain. Hence it is well suited to model gaits, sub-gaits and their relationships. It enables us to compactly represent a wide variety of knowledge though it impose some hard constraints [214].

Statistical Relational Learning (SRL) is a branch of *Artificial Intelligence* (AI) that is concerned with models of domains that exhibit both uncertainty (which can be dealt with using statistical methods) and complex, relational structure. Typically, the knowledge representation formalisms developed in SRL use two entities. The first one comprises of a subset of first-order logic, intends to describe relational properties of a domain in a general manner (universal quantification). The second one built upon probabilistic graphical models (such as Bayesian networks or Markov networks) aims to model the uncertainty. Significant contributions to the field have been made since the late 1990s. However, only very recently, Richardson and Domingos [214], members of the SRL group at the University of Washington, introduced Markov Logic Networks (MLNs). The objective of MLN is to soften the hard constraints imposed by first-order logic by combining logic and graphs. In First-order logic, formulas are perceived as hard constraints: a world (model) that violates even a single formula is impossible. On the other hand, in Markov logic, formulas are perceived as soft constraints: a world that violates a formula is less probable than the one which satisfies it, other things being equal, but not impossible.

The underlying problem of gait recognition demands a coherent framework which can integrate information from three domains, namely imaging, logic and graphs. In this chapter we will propose a basic three-layer architecture which can fuse these three domains to attack the *Gait Recognition (GR)* problem. We will initially use the imaging domain to represent gait motion in terms of silhouettes. Similar to the

PROCIM architecture, component based classification will be done at the imaging layer. Then we will define a first-order Knowledge-Base (KB) to represent and reason the classification results extracted from the imaging layer. We will hypothesize the proposed component-based GR paradigm in terms of simple inference rules. Further, the proposed framework strategically deploys MLN to learn gait component relationships defined in the KB. Finally gaits are classified using the proposed inference based rules.

The rest of the chapter is organized as follows. In Section 6.2 we present related work from the Statistical Relational Learning literature. Relevant background about first-order logic and MLN is provided in Section 6.3. Section 6.4 briefs the proposed 3-Layer fusion architecture of the proposed approach. We formulate the proposed Knowledge-Base, which serves as the backbone of the approach in Section 6.5. We provide experimental validations of the approach using a standard gait dataset in Section 6.6. Finally we conclude this chapter in Section 6.7.

6.2 Related Work

Traditionally, the field of AI research has fallen into two subfields: one that has been focusing on logical representations, and the other one that has been emphasizing statistical techniques [215]. Complexity issues have been handled by logical AI approaches such as symbolic parsing, logic programming, description logics, classical planning, rule induction and so on. Uncertainty modeling has been examined by statistical AI approaches such as Bayesian networks, hidden Markov models, Markov decision processes, statistical parsing, neural networks and so on. However, intelligent agents must be able to handle both complexity and uncertainty for real-world applications. Pioneering attempts to integrate logic and probability in AI commenced in late 1980s [216–218]. Later, several authors began using logic programs to compactly specify Bayesian networks, an approach known as knowledge-based model construction (Eg. [219]). Many approaches have been proposed in the recent years, including, Bayesian logic programs [220], relational dependency networks [221], probabilistic relational models [222], stochastic logic programs [223] and others. As these approaches typically combine probabilistic graphical models with a subset of first-order logic they can be quite complex. Unlike these approaches MLN utilizes the full expressiveness of first-order logic and graphical models without any restrictions [215].

The MLN based approach proposed by Tran and Davis [224] addresses the problem of event modeling and recognition in visual surveillance in unconstrained scenarios. The authors illustrate their approach in the context of monitoring a parking lot, with the goal of matching people to the vehicles they arrive and depart in. It has been shown that common sense knowledge, specific to the domain under consideration, can provide useful constraints to reduce uncertainties and ambiguities. Background subtraction, human detection and tracking techniques were first applied to identify and track object locations. The orientation and direction of each car is estimated using its corresponding foreground blob and parking lot layout. The Knowledge-Base proposed in the approach represents intuitions in the form of meaningful predicates (logical functions). A spatial predicate, for example, $inTrunkZone(C, H)$, is generated when the foot location of person, H, intersects significantly with the trunk zone of the car,C. Identity maintenance predicates are evaluated using the distance between color histograms of the two participating objects. Simple commonsense rules such as the following are formulated using first-order logic:

* If a person disappears, he/she enters a nearby car

* If a person opens the trunk of a car, he/she will (likely) enter that car

Though the proposed MLN-based approach has not been compared with other relevant models that deploys logic for event recognition (Eg. [225, 226]), it serves as a good example to illustrate the application of MLN towards uncertainty modeling.

Wu and Weld [227] have proposed a MLN based approach to address the problem of ontology refinement. As an outcome of their research they have developed an autonomous system called the Kylin Ontology Generator(KOG), capable of building rich ontologies by combining *Wikipedia* info-boxes with *WordNet*, a lexical database for the English language, using statistical-relational learning. The resulting ontology contains subsumption relations and schema mappings between info-box classes of Wikipedia. Additionally, it maps these classes to WordNet. The authors have shown that the resulting ontology may be used to enhance Wikipedia with improved query processing and other features.

Though MLN is a newly developing probabilistic logic paradigm, it is gaining momentum in the field of AI as it offers a unified solution to model uncertainty and complexity. Motivated by the spirit of the SRL literature, the fusion based approach presented in this chapter applies MLN to attack the GR problem.

6.3 Brief Backround of First-order Logic and MLN

For an in-depth coverage of first-order logic the reader is encouraged to read [140]. However, it is note worthy to brief some basic concepts relevant to MLN from [214] here. In AI, a Knowledge-Base, KB, technically represents a single large formula as the formulas in a KB are implicitly conjoined. A ground term is a term containing no variables. A ground atom or ground predicate is an atomic formula all of whose arguments are ground terms. A possible world assigns a truth value to each possible ground atom. A formula is satisfiable iff there exists at least one world in which it is true. The basic inference problem in first-order logic is to determine whether a knowledge base KB entails a formula F, i.e., if F is true in all worlds where KB is true (denoted by $KB \models F$). This is often done by refutation: KB entails F iff $KB \cup \neg\, F$ is unsatisfiable.

MLNs are basically undirected graphical models, being developed using SRL techniques to unify logic and probabilistic reasoning. Being a new technique, it is continuously undergoing developments by the SRL research group of Washington State University. Under the MLN paradigm, each first-order logic formula F_i is associated with a non-negative real-valued weight w_i. Prior to learning, every instantiation of F_i is given the same weight. Using machine learning procedures these weights are duly updated after the learning phase (weight learning). An undirected network, called a Markov Network, is constructed such that,

* Each of its nodes correspond to a ground atom x_k.

* If a subset of ground atoms $x_{\{i\}} = \{x_k\}$ are related to each other by a formula F_i, then a clique C_i over these variables is added to the network. A weight w_i is associated with C_i and a feature f_i is defined as follows:

$$f_i(x_{\{i\}}) = \begin{cases} 1 & \text{if } F_i(x_{\{i\}}) \text{ is true} \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

Thus first-order logic formulas in our KB serve as templates to construct the Markov Network. This network models the joint distribution of the set of all ground atoms, X, each of which is a binary variable. It provides a means for performing probabilistic inference using,

$$P(X = x) = \frac{1}{Z} exp(\sum_i -w_i f_i(x_{\{i\}})) \tag{6.2}$$

where Z, the normalizing factor, is defined as,

$$Z = \sum_{x \in X} exp(\sum_i -w_i f_i(x_{\{i\}})) \tag{6.3}$$

If $\Phi_i(x_{\{i\}})$ is the potential function defined over a clique C_i, then

$$log(\Phi_i(x_{\{i\}})) = w_i f_i(x_{\{i\}}) \tag{6.4}$$

An MLN is obtained by attaching weights to the formulas (or clauses) in a first-order knowledge base, and can be viewed as a template for constructing Markov networks. Empirically several algorithms for MLN weight learning have been compared in terms of learning rates by Lowd and Domingos [228]. Each possible grounding of a formula in the KB yields a feature in the constructed network. Inference is performed by grounding the minimal subset of the network required for answering the query and running a Gibbs sampler over this subnetwork, with initial states found by the MaxWalkSat algorithm. Weights are learned by optimizing a pseudo-likelihood measure using the L-BFGS algorithm, and clauses are learned using the CLAUDIEN system. For a given probe, by observing the gaits triggered by its various components, we aim to query the most probable gallery instance with the aid of the learnt potentials readily available.

6.4 Proposed 3-Layer Architecture that Fuses Imaging, Logic and Probabilistic Graphical Domains

We abbreviate the proposed **G**ait **R**ecognition **M**odel which uses **M**arkov **L**ogic **N**etwork as (**GRM-MLN**). The 3-layer architecture employed by the GRM-MLN is shown in Figure 6.1. We briefly describe the three stages inherent in the proposed framework as follows:

i. Image Processing Layer (IPL)

   Initial image processing and component based classifications are performed at the image processing layer. Raw binary silhouettes which form the core input of the model are initially decomposed into three sub-gaits namely Left-Gait, Right-Gait and the Lower-Gait using basic image segmentation techniques after normalizing them using standard image processing techniques. The segmentation scheme proposed in Section 5.2 for the PROCIM architecture has been used here as well. GRM-MLN is a generic object recognition model and hence it can flexibly fit into any feature projection technique such as PCA or SVM. For demonstration sake we represent the silhouettes in terms of multi-linear tensors which has been recently applied by Lu et al. [196]. Eigen-tensor based features [196] are extracted from these gait components from which component based recognition is performed.

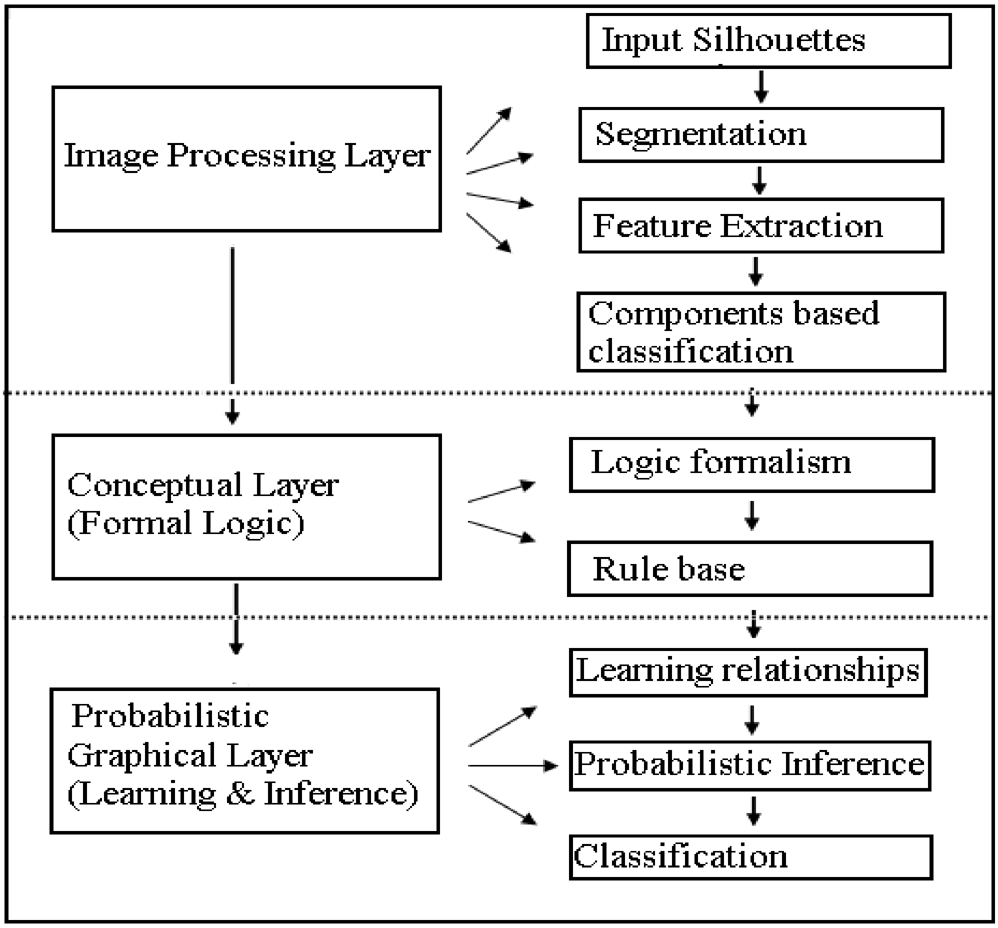


FIGURE 6.1: Three-Layer architecture of the proposed GRM-MLN

ii. Conceptual Layer (CL)

This layer is comprises of a set of predicate definitions (first-order formulas) and a Knowledge-Base (KB) that governs how various gait components can be relatively combined. The information gained from the weak classifiers based on components based recognition from the IPL is transformed into the logical layer in terms of ground atoms.

iii. Probabilistic Graphical Layer (PGL)

Each of the gait components individually and/or collectively could contribute to the recognition of a subject. Undirected graphs employed by the GRM-MLN represent the dependencies between the components and the gaits which they recognize. The graph has a node for each variable, and the model has a potential function (weight) for each clique in the graph. From the training dataset the potentials encoded by the graphs receive possible groundings for the atomic formulas which are precisely governed by the rules in the KB of the logical layer. This enables GRM-MLN to learn characteristics about gait components and their relationships. A given probe is subjected to segmentation and component based classification is performed. The information obtained from these weaker classifiers are fed into the GRM-MLN as evidences and finally the most probable gait recognized for the given probe is determined.

6.5 Formulation of the proposed Knowledge Base

We intend to represent the knowledge about the gait domain which we perceive as, gaits, potential sub-gaits and the various relationships between them. As the gait domain comprises real world object entities and the behavior that governs these entities, it is appropriate to use first-order logic to represent the gait domain. A first-order knowledge base, KB, is a set of sentences or formulas in first-order logic [229]. Formulas are constructed using four types of symbols: constants, variables, functions, and predicates. Constant symbols represent objects in the gait domain. For example $G1$ and $G5$ refers to the gait of the first and fifth subject. Variable symbols range over the objects in the domain. For example *people* is a variable which can range over $G1$, $G2$ and so on. Function symbols are used to represent particular behavior of an object or a set of objects. For example, $LowerLeft$ is a function used to represent the left leg motion of a gait. Predicate symbols represent relations among objects in the domain. For example the predicate $LowerGait(person)$ relates a person with respect to his particular gait behavior. An atomic sentence is an indivisible formula, represented by a single proposition symbol (Eg. $\neg P$), which stands for a proposition that can be either true or false. A KB consists of a set of formulas which are constructed from atomic sentences. An efficient KB is formed by few predicates and a compact rule-base, where each rule is stated clearly and concisely [140]. Keeping this as a guideline, we construct our gait KB using the following three evidence predicates: *LftGait(person), RgtGait(person) and LowerGait(person)*. These three evidence predicates represent the binary states of the three potential sub-gaits which we identified in Section 5.5.2. Each sub-gait either individually or collectively might trigger the recognition of a subject which in turn leads to a series of component based recognition rules. These rules that characterize the recognition potential of each of the above introduced predicates are defined as follows:

$$LftGait(person) \implies Recognize(person) \tag{6.5}$$

$$RgtGait(person) \implies Recognize(person) \tag{6.6}$$

$$LowerGait(person) \implies Recognize(person) \tag{6.7}$$

Logically the following rules are derived from the above rules using conjunction.

$$LftGait(person) \wedge RgtGait(person) \implies Recognize(person) \tag{6.8}$$

$$LftGait(person) \wedge LowerGait(person) \implies Recognize(person) \tag{6.9}$$

$$RgtGait(person) \wedge LowerGait(person) \implies Recognize(person) \tag{6.10}$$

Subtle gait motions can be derived by the conjunction (intersection) of sub-gaits. For example, the conjunction of $LftGait$ and $LowerGait$ of a person yields subtle lower-left-leg motion of the person. Hence, intuitively, equations 6.9 and 6.10, help GRM-MLN to infer, how well do the subtle sub-gait motions viz., lower-left-leg and lower-right-leg, respectively, contribute to the overall gait recognition process. Physically modeling these subtle sub-gait motions, obviously involves incorporation of sophisticated segmentation algorithms at the cost of additional computational costs. GRM-MLN takes advantage of logic to model such intrinsic sub-gait patterns without the need of any sophisticated segmentation techniques. By way of establishing simple logical relationships, the proposed GRM-MLN model intend to learn subtle relationships from the gait domain by applying Markov-Logic Networks.

For our KB, we have mostly used two of the following simple inference based rules.

- Modus Ponens which can be written as follows:

$$\frac{\alpha \implies \beta, \quad \alpha}{\beta} \tag{6.11}$$

  This means that, whenever any sentences of the form $\alpha \implies \beta$ and α are given, then the sentence β can be inferred. For example, if GRM-MLN knows that

$$LftGait(G7) \wedge RgtGait(G7) \implies Recognize(G7), \tag{6.12}$$

  that is the gait of the subject $G7$ is influenced by both his left-gait and right-gait motions, then, whenever $LftGait(G7) \wedge RgtGait(G7)$ has been observed, then the recognition state of $G7$, $Recognize(G7)$, can be inferred (queried).

- And-Elimination which is another useful inference rule can be written as follows:

$$\frac{\alpha \wedge \beta}{\alpha} \tag{6.13}$$

  which says that, from a conjunction, any of the conjuncts can be inferred. For example, from the knowledge of $LftGait(G7) \wedge RgtGait(G7)$, $RgtGait(G7)$ can be inferred. Predicates which are used for inference are called query predicates (Eg. $Recognize(person)$).

6.5.1 Representing Rules in Conjunctive Normal Form

Considering that the percepts of the GRM-MLN relies on the evidence perceived by the three predicates, we hypothesize that: *"A person is recognized, if and only if* *atleast one of his/her three sub-gaits, contributes to the overall recognition process of the person"*. Logically this hypothesis can be formulated as:

$$LftGait(person) \vee RgtGait(person) \vee LowerGait(person) \iff Recognize(person) \tag{6.14}$$

For automated inference, it is often convenient to convert formulas to a more standard form, called *conjunctive normal form (CNF)* [214]. A KB in CNF form which is otherwise known as clausal form is a conjunction of clauses, a clause being a disjunction of literals. Every KB in First-order logic can be converted to clausal form using a mechanical sequence of steps. A formula is satisfiable if it is true in some model. If a sentence, α, is true in a model m, then we say than m satisfies α. In other words m is a model of α. There are special algorithms (Eg. WalkSAT [230]) to solve satisfiability (SAT) of a formula very efficiently if the formula is written in a CNF form. The set of rules that we have defined, that is the rule-base, and their corresponding CNFs are given in Table 6.1

TABLE 6.1: List of rules of the proposed knowledge-base and their corresponding CNFs (conjunctive normal form)

Rules used in the Knowledge-base	**Clausal form (CNF) of the rules**
LftGait(person) $\implies$ Recognize(person)	$\neg$ LftGait(person) $\vee$ Recognize(person)
RgtGait(person) $\implies$ Recognize(person)	$\neg$ RgtGait(person) $\vee$ Recognize(person)
LowerGait(person) $\implies$ Recognize(person)	$\neg$ LowerGait(person) $\vee$ Recognize(person)
LftGait(person) $\wedge$ RgtGait(person) $\implies$ Recognize(person)	$\neg$ LftGait(person) $\vee$ $\neg$ RgtGait(person) $\vee$ Recognize(person)
LftGait(person) $\wedge$ LowerGait(person) $\implies$ Recognize(person)	$\neg$ LftGait(person) $\vee$ $\neg$ LowerGait(person) $\vee$ Recognize(person)
RgtGait(person) $\wedge$ LowerGait(person) $\implies$ Recognize(person)	$\neg$ RgtGait(person) $\vee$ $\neg$ LowerGait(person) $\vee$ Recognize(person)
LftGait(person) $\vee$ RgtGait(person) $\vee$ LowerGait(person) $\iff$ Recognize(person)	($\neg$ LftGait(person) $\vee$ Recognize(person)) $\wedge$ ($\neg$ RgtGait(person) $\vee$ Recognize(person)) $\wedge$ ($\neg$ LowerGait(person) $\vee$ Recognize(person)) $\wedge$ $\wedge$ ($\neg$ Recognize(person) $\vee$ LftGait(person)) $\wedge$ ($\neg$ Recognize(person) $\vee$ RgtGait(person)) $\wedge$ ($\neg$ Recognize(person) $\vee$ LowerGait(person))

The CNFs shown in Table 6.1 are derived by applying a set of standard logical equivalence relations [231] which are given in Table 6.2.

TABLE 6.2: Logical equivalence relations of First-Order Logic

Standard logical equivalence relations	**In words**
$(A \wedge B) \equiv (B \wedge C)$	commutativity of $\wedge$
$(A \vee B) \equiv (B \vee C)$	commutativity of $\vee$
$((A \wedge B) \wedge C) \equiv (A \wedge (B \wedge C))$	associativity of $\vee$
$((A \vee B) \vee C) \equiv (A \vee (B \vee C))$	associativity of $\vee$
$\neg(\neg A) \equiv A$	double-negation elimination
$(A \implies B) \equiv (\neg B \implies \neg A)$	contraposition
$(A \implies B) \equiv (\neg A \vee B)$	implication elimination
$(A \iff B) \equiv (A \implies B) \wedge (B \implies A)$	biconditional elimination
$\neg(A \wedge B) \equiv (\neg A \vee \neg B)$	De Morgan's Law
$\neg(A \vee B) \equiv (\neg A \wedge \neg B)$	De Morgan's Law
$(A \wedge (B \vee C)) \equiv ((A \wedge B) \vee (A \wedge C))$	distributivity of $\wedge$ over $\vee$
$(A \vee (B \wedge C)) \equiv ((A \vee B) \wedge (A \vee C))$	distributivity of $\vee$ over $\wedge$

6.5.1.1 Proof of Hypothesis

Similar to how algebraic identities are applied to derive algebraic formulas, we will apply these logical relations to derive the proof of the CNF of our hypothesis represented by equation 6.14, last row of Table 6.1, a typical biconditional rule.

Applying biconditional elimination, equation 6.14 transforms to the conjuntion of eqs. 6.15 and 6.16.

$$LftGait(person) \vee RgtGait(person) \vee LowerGait(person) \implies Recognize(person) \tag{6.15}$$

$$Recognize(person) \implies LeftGait(person) \vee RgtGait(person) \vee LowerGait(person) \tag{6.16}$$

Applying implication elimination to eq. 6.15

$$\neg(LftGait(person) \vee RgtGait(person) \vee LowerGait(person)) \vee Recognize(person) \tag{6.17}$$

Applying De Morgan's Law to eq. 6.17

$$(\neg LftGait(person) \wedge \neg RgtGait(person) \wedge \neg LowerGait(person)) \vee Recognize(person) \tag{6.18}$$

$$\begin{aligned}(\neg LftGait(person) \vee Recognize(person)) \wedge (\neg RgtGait(person) \vee Recognize(person)) \\ \wedge (\neg LowerGait(person) \vee Recognize(person))\end{aligned} \tag{6.19}$$

Applying implication elimination to eq. 6.16

$$\neg Recognize(person) \vee (LftGait(person) \wedge RgtGait(person) \wedge LowerGait(person)) \tag{6.20}$$

Applying distributivity of ∨ over ∧ to eq. 6.20

$$\begin{aligned}(\neg Recognize(person) \vee LftGait(person)) \wedge (\neg Recognize(person) \vee RgtGait(person)) \\ \wedge (\neg Recognize(person) \vee LowerGait(person))\end{aligned} \tag{6.21}$$

The conjunction of eqs. 6.19 and 6.21 represent the CNF of eq. 6.14 as these are the corresponding CNFs of eqs. 6.15 and 6.16 which are in turn derived from eq. 6.14.

As described above, all the possible rules (clauses) have been defined using first-order logic and a precise Knowledge Base (KB) of gait components and their relationships has been formed. Each rule defined in the KB corresponds to the event of a component based recognition mechanism. Initially at the Image Processing Layer (IPL), eigen-tensor based feature space has been constructed for the various gait components from the training samples. By projecting the eigen-tensor feature of a gait component of a test sample over this feature space, recognition has been performed and the results are transformed into the Conceptual Layer (CL) in terms of ground atoms. In other words, in the event of a component or set of components have influenced the recognition of a gait, the corresponding rule in the KB will receive a grounding.

6.6 Experiments and discussion

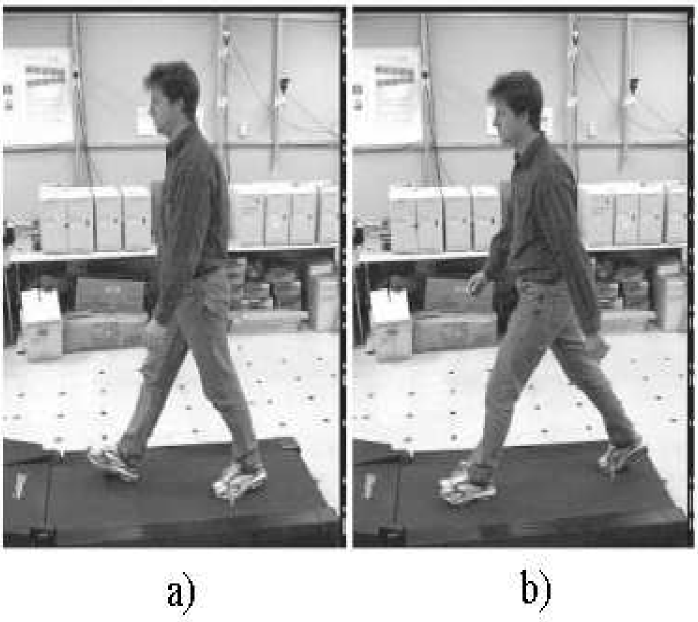


Figure 6.2: Typical samples from the CMU gait database showing the gait of a subject walking on a treadmill set in the middle of a room for the conditions of a) slow walk and b) fast walk.

We have used the CMU Mobo data set [5] which consists of gait sequences of subjects walking on a treadmill, positioned in the middle of a room. The dataset was developed by Collins et al. of Carnegie Mellon University(CMU), who also proposed the CMU gait algorithm which has been briefed in section 4.3.1.2. The dataset contains six simultaneous motion sequences of 25 subjects (23 male, 2 female) walking on a treadmill. The 3CCD progressive scan images have a resolution of 640x480. Each subject is recorded performing four different types of walking: slow walk, fast walk, inclined walk, and slow walk holding a ball (to inhibit arm swing). For our experiments we have used gait silhouettes representing the slow walk and fast walk video sequences. More than 8000 images are captured per subject. Sample video frames for a typical subject for two typical conditions viz., slow walk and speed walk are shown in Figure 6.2. Each sequence is 11 seconds long, recorded at 30 frames per second. For training and testing we have used the slow walk and fast walk sequences respectively. The average walking speed of the treadmill was set to 2.06 miles per hour (mph) for capturing the slow walk gait sequences. For the fast walk this was set as 2.82 mph. The speed of the treadmill was adjusted to be at a comfortable walking speed for the subjects for both the slow walk and fast walk.

We have implemented GRM-MLN using an open source software called "Alchemy" offered by the *Statistical Machine Learning Group, University of Washington, http://alchemy.cs.washington.edu*, [214]. MLN weight learning has been performed to learn the potentials of each of the formulas in the knowledge base using the training samples. The normalized weights which represent the recognition potential of various sub-gaits and their logical relationships are shown in Figure 6.3.

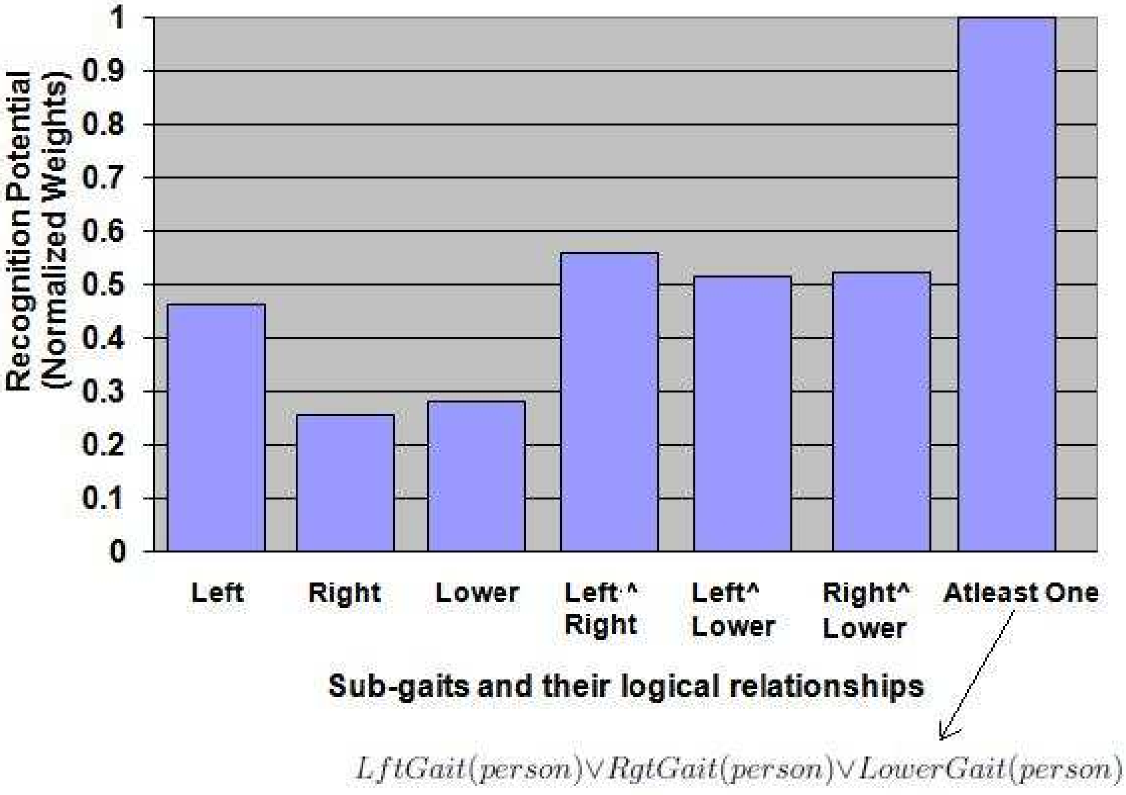


FIGURE 6.3: Learnt potentials of evidence predicates for training samples comprising slow walk

It can be seen that various gait components such as left, right and lower body motion contribute around 25% to 45% towards the overall gait recognition. Lower body motion without any further component based interpretation contributes to a recognition potential of about 28%. However the efficient fusion of lower body motion with left and right gait symmetries have enabled GRM-MLN to learn the recognition potentials of two vital components viz., lower-left and lower-right gait components. The average recognition potential of these components contributes to around 52% which is considerably better than the lower body motion alone. Similarly the fusion of left and right gaits have yielded a recognition potential of about 56%. This is considerably more when compared to their individual contribution. This illustrates a significant advantage of the fusion based mechanism deployed by GRM-MLN.

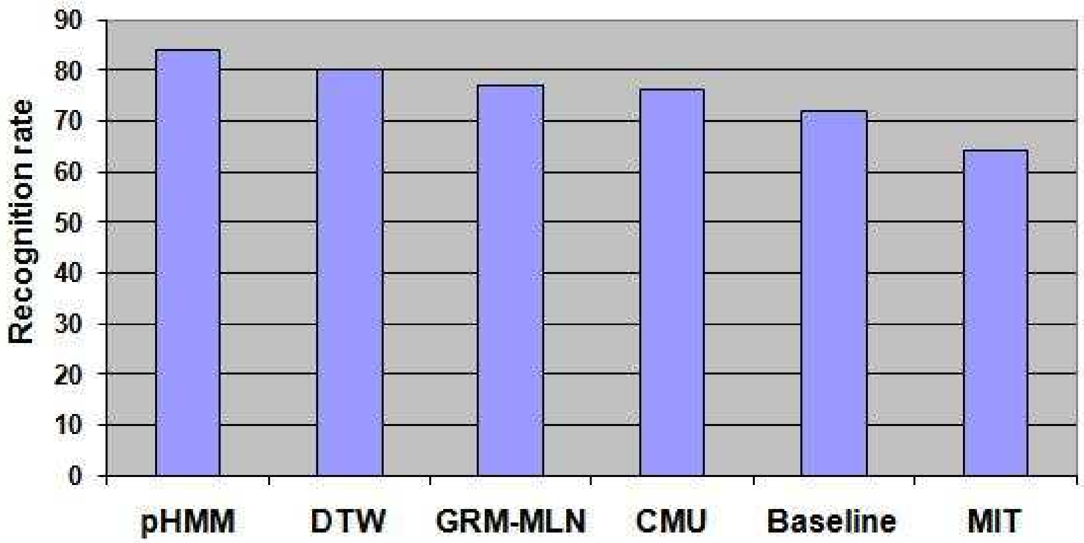


Figure 6.4: Comparison of GRM-MLN with GDN [9],UMD [10],CMU [11], Baseline [12] and MIT [13]

Further we have compared the recognition rates of GRM-MLN with state-of-the-art gait recognizers with respect to the CMU dataset. From the bar-chart shown in Figure 6.4, it can be seen that the proposed GRM-MLN algorithm competes well with other standard algorithms. We have briefed about the pHMM and DTW algorithms in section 4.3.1.1. We have seen that the pHMM algorithm proposed by Liu and Sarkar [9] relies on manually created silhouettes. The DTW algorithm proposed by Veeraraghavan and Chowdhury [10] derives geometric information of the walking person from several landmark points which are manually marked on the gait video. Though the recognition rate of GRM-MLN is relatively lower than pHMM and DTW, it has the advantage of avoiding such manual interventions.

6.7 Summary

We have proposed a simple but yet efficient statistical relational learning technique to reason and recognize gaits. This study shows how a simple component based gait reasoning approach can be coherently modeled using Markov Logic Networks. The proposed GRM-MLN has a natural generative semantics, which can establish dependencies between gait components and exploit these dependencies to successfully classify gaits. For a newly emerging biometric like gait, every piece of contribution such as GRM-MLN would be a milestone. But as the newly developing SRL-based MLN is still undergoing developments, it has several practical limitations when compared to the well established *Bayesian Network toolbox*. For example, it is not possible in MLNs to define a potential function that depends on certain operations (eg. dot product) between two object entities [232]. The reason is that each potential function is a (learnt) constant. The root problem in MLN is that it has not moved beyond weighted first-order logic to more general weighted algebraic constraint systems. Hence directly comparing the Bayesian Network-based PROCIM architecture with GRM-MLN is not appropriate without modeling key psychophysical principles such as *influence strengths*. For future avenues, we intend to formulate advanced concepts such as *influence strengths*, which has been proposed in Chapter 5 for the PROCIM architecture, using MLNs or other graphical models such as factor graphs. However the basic framework presented in this chapter provides us a good starting point to explore more on statistical relational learning (SRL) concepts and MLNs for our future research.

Chapter 7

Conclusions and Future Work

This thesis has investigated two potential biometrics namely face and gait recognition under unconstrained scenarios. Psychophysically feasible novel probabilistic models have been proposed based on recent computer vision techniques. This chapter will present the overall conclusions derived for face and gait recognition in section 7.1 and 7.1 respectively. Finally future research avenues will be discussed in section 7.3.

7.1 Face Recognition

As an outcome of Chapter 2, we identified several key psychophysical principles that govern humans to recognize faces under unconstrained scenarios, where complexities such as major occlusions, noise, illumination variation, scale and so on impose vast uncertainty to the recognition process. Inspired by these psychophysical principles, Chapter 3 defined a phenomenon called *similarity mappings* and proposed a novel **PRO**babilistic **C**omponent **I**nterpretation **M**odel (**PROCIM**) based on this phenomenon. PROCIM deployed Bayesian Networks(BN) to scientifically model this intuitive similarity mappings. Importantly, it provided sound visual means using graphs to analyse and interpret how the information derived from various subsamples can collectively contribute to tackle uncertainties. Further the impact of varied degrees of occlusions over similarity mappings has been clearly demonstrated. It has been justified that even some of the popular occlusion models failed to exhibit such visual capabilities. The transparency exhibited by PROCIM is due to the intuitive psychophysical nature of the model.

Further a novel physical property called *Influence Strength*, Z, has been defined as an outcome of Chapter 3. It has been hypothesized that *"The face being recognized by observing a subsample of an occluded probe face will be more similar to the probe, if Z's magnitude is high"*. A novel formula to make decision under uncertainty has been proposed by effectively unifying probability theory and the crucial influence strengths.

Extensive experimental validations have been presented to compare PROCIM with state-of-the-art occlusion models. Against real occlusions such as sunglasses and scarf, PROCIM reported promising recognition rates of about 90% within 7 ranks, with respect to the AR dataset. Even the very recently proposed AWSGA [126] algorithm is unable to achieve this performance within 20 ranks. But AWSGA has the advantage of using one training sample per class whereas PROCIM uses four training samples per class.

A novel evaluation method called *Discrete Random Occlusion Test (DROT)* has been proposed to simulate realistic occlusions. Empirical evaluations have been performed, using both CROT (Continuous Random Occlusion Test) and DROT, using the defacto standard DARPA's FERET dataset. With respect to conventional tests, PROCIM yielded recognition rates of 94.3% and 90.1% in the presence of moderate and major occlusions respectively, within the top three ranks. When PROCIM and the classical PCA were subjected to the DROT, they yielded overall recognition rates of 82.7% and 50.3% within the top three ranks. The wide performance gap between the two approaches justifies the advancement of PROCIM over the conventional PCA.

One of the limitations of PROCIM is that it needs more training samples. However, the fact that PROCIM has the ability to converge to peak performance within a few top ranks, indicates that PROCIM promises to recognize the actual subject, in unconstrained scenarios, reasonably well. If a biometric enabled security system can provide such an ability, it will give the criminal investigation team a considerable advantage, which is a significant advancement in the field of biometrics.

7.2 Gait Recognition

The literature review presented in chapter 4 concluded that Gait Recognition (GR) being a newly emerging biometric need to learn lessons from other matured biometrics such as face recognition. It has been found that silhouette-based GR techniques are gaining momentum among computer vision researchers. Recent studies have concluded that silhouette shape has equal, if not more, recognition potential than gait kinematics. Based on this inspiration two silhouette-based GR approaches have been presented in Chapters 5 and 6.

As an outcome of chapter 5, it has been discovered that, to perform GR, segmentation of specific body components such as head, arms, torso and legs are not required. Obviously such accurate segmentation requires manual labour. We have shown that PROCIM could derive useful information from various body components using a simple automatic segmentation strategy. To achieve this, a set of novel gait components called sub-gaits have been formulated based on the intuition that for some individuals certain gait aspects, e.g. lower gait motion, might be more discriminative than say upper gait motion. For others, left or right gait motion could trigger their identification. The extended PROCIM architecture proposed in chapter 5, enabled such an analysis to be factored into the recognition process using Bayesian Networks. Further the proposed novel modular training scheme, enabled PROCIM to represent and learn subtle walking patterns of human gaits. It has been demonstrated that though the silhouettes have been inherently noisy and some of them contained missing heads, arms, torso and legs, PROSIM has been robust enough to tackle such uncertainties.

Empirical results proved that the sub-gaits viz., lower-gait, left-gait and right-gait exhibit higher recognition potential than upper-gait and mid-gait. With respect to the DARPA's humanID challenge dataset, two experimental settings have been used viz., PROCIM-a (limited gallery set) and PROCIM-b (limited probe set). For the rank-1 performance, PROCIM-a yielded about 90%, 24% and 51% recognition rates for the most easy experiment, the most difficult experiment and the mean of all experiments respectively. For these typical categories of experiments, considering the rank-5 performance , recognition rates of 93%, 72% and 75% have been achieved. On the other hand, with respect to the rank-1 performance, PROCIM-b yielded about 93%, 72% and 78% recognition rates for the most easy experiment, the most difficult experiment and the mean of all experiments respectively. For the same category of experiments, with respect to rank-5 performance, it achieved recognition rates of 98%, 88% and 90%. Eventually, the performance of PROCIM-b is better than PROCIM-a. The reason is that it used multiple gallery sets and hence gained more training experience than PROCIM-b. PROCIM-b outperformed other state-of-the-art algorithms in majority of the DARPA experiments which justifies its robustness.

Further experiments have been performed with the CASIA dataset to prove the generalization capacity of PROCIM. When tested by varying the carrying condition alone (i.e. for the same view), PROCIM and GEI yielded recognition rates of about 87% and 68% respectively. When tested by varying the clothing condition alone, PROCIM and GEI yielded recognition rates of about 50% and 29% respectively. This indicates that clothing is a tough test as the occlusion caused by long coat (most of the body parts are occluded by a long coat) imposes vast uncertainty to the recognition process. For a small viewing variation of 18°, PROCIM and GEI yielded recognition rates of about 49% and 39% respectively. However when viewing is varied extremely (trained with 0° and tested with 90°) coupled with clothing variation, the recognition rates have been considerably degraded to

8.3% and 2.5% respectively, by PROCIM and GEI. We see that PROCIM shows significant improvement in performance over cloting and carrying conditions.

The GRM-MLN algorithm proposed in Chapter 6 reveals that subtle gait motion can be coherently modeled by fusing diverse domains viz., imaging, logic and graphs. It has been demonstrated that GRM-MLN can establish dependencies between gait components (sub-gaits) and exploit these dependencies to successfully classify gaits. It can be seen that various gait components such as left, right and lower body motion contribute around 25% to 45% towards the overall gait recognition. Lower body motion without any further component based interpretation contributes to a recognition potential of about 28%. However the efficient fusion of lower body motion with left and right sub-gaits have enabled GRM-MLN to learn the recognition potentials of two subtle sub-gait motions viz., lower-left and lower-right gait. The average recognition potential of these components contributes to around 52% which is considerably better than the lower body motion alone. Similarly the fusion of left and right sub-gaits have yielded a recognition potential of about 56%. This is considerably more when compared to their individual contribution. This illustrates a significant advantage of the fusion based mechanism deployed by GRM-MLN.

The proposed GRM-MLN algorithm competes well with other state-of-the-art algorithms. Algorithms such as pHMM and DTW relies on manually created silhouettes and manual labeling. Though the recognition rate of GRM-MLN is relatively lower than pHMM and DTW, it has the advantage of avoiding such manual interventions.

There are some limitations with the newly developing SRL-based MLNs. For example, it is not possible in MLNs to define a potential function that depends on certain operations (eg. dot product) between two object entities [232]. The reason is that each potential function is a (learnt) constant. The main drawback of MLN is that, it has not moved beyond weighted first-order logic to more general weighted algebraic constraint systems.

7.3 Future Work

One of the interesting avenues for future directions could be addressing: “The proposed probabilistic models do not have direct dependencies among object components and does this detract from the power of the modeling?”. We have applied Bayesian Networks (which use directed edges) in the proposed framework to exploit the conditional independence properties that exists between objects and their components to achieve robust gait recognition. Such independence assumptions reduce the number of parameters in the model, and therefore making the model computationally feasible for real time applications. However setting dependencies among parts could be modeled using undirected links. Graphical models such as Markov networks [213] which use undirected graphs can be employed to capture dependency among various components. In this regard, it will be an interesting avenue in the future to apply undirected graphical models, to investigate the impact of dependencies between object components and ultimately how they would influence the whole recognition process.

We have modeled *influence strengths* under the Bayesian network framework. Alternatively whether the first-order logic oriented MLNs can be used to model this concept, needs further investigation. Further, we have explicitly modeled a novel robust formula which has enabled PROCIM to make meaningful decision under uncertainty by bridging probability theory and utility theory. This can be alternatively modeled using decision networks (influence diagrams), a generalization of a Bayesian network model. However, this needs more understanding of utility theory. So for our future assignments we will consider exploring these complimentary methods.

Face and gait biometrics serve as natural candidates to be fused to result in multimodal biometrics, as both have the advantage of being non-invasive. At a distance, gait can be used and gradually when the individual approaches, face images could provide additional cues. Hence at a near distance, they may generally be fused to advance the recognition accuracy. This challenging problem of biometric fusion would bring new dimensions to future research and development opportunities.

Having seen that PROCIM has the flexibility and robustness to address both face and gait recognition problems, for future avenues we would like to extend it to address other object recognition problems such as fingerprint recognition. Owing to the fact that fingerprint technology is in forensic practice for more than a century, there is a popular misconception that it is a fully solved problem. Latent fingerprints (taken from a crime scene) may exhibit only a small portion of the surface of the finger and may be prone to uncertainty factors such as smudges, distortions, or both, depending on how they were deposited. PROCIM, being an uncertainty model, might provide a reasonable solution to such intricate issues. The challenge lies in finding the optimal discriminative subsamples within the fingerprint domain.

The Thales Group, a leading defence industry and a couple of organizations from Malaysia have shown interest on our award winning probabilistic models. We will explore commercialization possibilities by having suitable knowledge transfer agreements with such organizations. An exciting future assignment will be acquiring the knowledge and expertise that are vital to convert the proposed probabilistic models to DSP chips. This would pave us the way to expose our technology from our intelligent systems lab to the global consumer electronics industry market.

Bibliography

[1] http://www.biometricgroup.com. International Biometric Group, 2009.

[2] P.J.Phillips, H.Moon, and S.A.Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000.

[3] M. P. Murray, A. B. Drought, and R. C. Kory. Walking patterns of normal men. *Journal of Bone and Joint Surgery*, 46:335–360, 1964.

[4] J. E. Cutting and L. T. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. Technical report, 1977.

[5] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition? In *Third workshop on empirical evaluation methods in computer vision*, 2001.

[6] A. Kale, N. Cuntoor, B. Yegnanarayana, A. N. Rajagopalan, and R. Chellappa. Gait analysis for human identification. *Audio-and Video-Based Biometric Person Authentication, Proceedings*, 2688:706–714, 2003.

[7] Y. X. Liu, R. Collins, and Y. H. Tsin. Gait sequence analysis using frieze patterns. *Computer Vision - Eccv 2002, Pt II*, 2351:657–671, 2002.

[8] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos. An angular transform of gait sequences for gait assisted recognition. *Icip: 2004 International Conference on Image Processing, Vols 1- 5*, pages 857–860, 2004.

[9] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:863–876, 2006.

[10] A. Veeraraghavan, A. R. Chowdhury, and R. Chellappa. Matching shape sequences in video with applications in human movement analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27:1896–1909, 2005.

[11] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 366–371, 2002.

[12] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K.W.Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:162–177, 2005.

[13] L. Lee. *Gait Analysis for Classification*. PhD thesis, Massachusets Inst. of Technology, 2003.

[14] E. Aarts and S. Marzano. *The New Everyday: Visions of Ambient Intelligence*. 010 Publishers, 2003.

[15] http://www.boozallen.com/media. Biometrics-Enabled Intelligence, February 2010.

[16] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[17] B. V. K. V. Kumar, M. Savvides, and C. Xie. Correlation pattern recognition for face recognition. *Proc. of IEEE special issue on Biometrics: Algorithms & Applications*, 94:1963 – 1976, 2006.

[18] The Oxford English Dictionary, 1951.

[19] Aristotle. On the Motion of Animals, B.C.350, 2004.

[20] M.S. Nixon and J.N. Carter. Automatic Gait Recognition for Human ID at a Distance. Technical Report N68171-01-C-9002, University of Southampton, Nov. 2004.

[21] D. Winter. *The Biomechanics and Motor Control of Human Gait, 2nd Ed., Waterloo, 1991.* Waterloo, 1991.

[22] E.T. Jaynes. *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.

[23] C.M. Bishop. *Pattern recognition and machine learning.* Springer, 2006.

[24] D. Koller and N. Friedman. *Probabilisitc graphical models.* The MIT Press, Cambridge, 2009.

[25] F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs.* Springer, 2007.

[26] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[27] B. Moghaddam, C.Nastar, and A.Pentland. A Bayesian similarity measure for deformable image matching. *Image & Vision Computing*, 19:235–244, 2001.

[28] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans:nineteen results all computer vision researchers should know about. *Proceedings of the IEEE special issue on biometrics*, 94(11):1948–1962, 2006.

[29] A. Fific. *Emerging holistic properties at face value: assessing characteristics of face perception.* PhD thesis, Dept. of Psychology and Cognitive Science, Indiana University, 2005.

[30] Y. Zana, R. M. Cesar, R. Feris, and M. Turk. Local approach for face verification in polar frequency domain. *Image and Vision Computing*, 24(8): 904–913, 2006.

[31] M.S.Bartlett and J.R.Movellan. Face recognition by independent component analysis. *IEEE Trans. on Neural Networks*, 13(6):1450–1464, 2002.

[32] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja. Memory-based face recognition for visitor identification. In *IEEE Intl. Conference on Automatic Face and Gesture Recognition*, pages 214–220, 2000.

[33] L. Wiskott, J. M. Fellous, N. Kruger, and C. vonderMalsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[34] I.S. Bruner and R. Tagiuri. *The perception of people*, volume 2. Addison-Wesley, 1954.

[35] V. Bruce. *Recognizing faces.* Lawrence Erlbaum Associates, 1988.

[36] H. Ellis, J. Shepherd, and G. Davies. Investigation of use of photo-fit technique for recalling faces. *British Journal of Psychology*, 66(FEB):29–37, 1975.

[37] 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*, 14(8):1724–1733, 1997.

[38] P. J. B. Hancock, V. Bruce, and M. A. Burton. A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research*, 38(15-16):2277–2288, 1998.

[39] P. Kalocsai, W. Y. Zhao, and E. Elagin. Face similarity space as perceived by humans and artificial systems. *Automatic Face and Gesture Recognition - Third IEEE International Conference Proceedings*, pages 177–180, 1998.

[40] M. Behrmann, R. S. Zemel, and M.C.Mozer. Object-based attention and occlusion. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1011–1036, 1998.

[41] B. J. Frey and N. Jojic. A comparison of algorithms for inference and learning in probabilistic graphical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1392–1416, 2005.

[42] J. Lin, J. Ming, and D. Crookes. Robust face recognition using posterior union model based neural networks. *IET Computer Vision*, 3(3):130–142, 2009.

[43] V. Bruce, Z. Henderson, K. Greenwood, P. J. B. Hancock, A. M. Burton, and P. Miller. Verification of face identities from images captured on video. *Journal of Experimental Psychology-Applied*, 5(4):339–360, 1999.

[44] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce. Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, 10 (3):243–248, 1999.

[45] C. H. Liu, H. Seetzen, A. M. Burton, and A. Chaudhuri. Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology-Applied*, 9(1):33–41, 2003.

[46] D. A. Roark, A. J. O'Toole, and H. Abdi. Human recognition of familiar and unfamiliar people in naturalistic video. *IEEE International Workshop on Analysis and Modeling of Face and Gestures*, pages 36–41, 2003.

[47] T. Tieger and L. Ganz. Recognition of faces in the presence of 2-dimensional sinusoidal masks. *Perception and Psychophysics*, 26(2):163–167, 1979.

[48] A. Fiorentini, L. Maffei, and G. Sandini. The role of high spatial-frequencies in face perception. *Perception*, 12(2):195–201, 1983.

[49] T. Hayes, M. C. Morrone, and D. C. Burr. Recognition of positive and negative bandpass-filtered images. *Perception*, 15(5):595–602, 1986.

[50] N. P. Costen, D. M. Parker, and I. Craw. Spatial content and spatial quantization effects in face recognition. *Perception*, 23(2):129–146, 1994.

[51] E. Peli, E. Lee, C. L. Trempe, and S. Buzney. Image enhancement for the visually impaired, the effects of enhancement on face recognition. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 11(7): 1929–1939, 1994.

[52] R. Nasanen. Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39(23):3824–3833, 1999.

[53] H. Ojanpaa and R. Nasanen. Utilisation of spatial frequency information in face search. *Vision Research*, 43(24):2505–2515, 2003.

[54] S. C. Dakin and R. J. Watt. Biological bar codes in human faces. *Journal of Vision*, 9(4):1–10, 2009.

[55] M. S. Keil. Does face image statistics predict a preferred spatial frequency for human face processing? *Proceedings of the Royal Society B-Biological Sciences*, 275(1647):2095–2100, 2008.

[56] G. J. Burton and I. R. Moorhead. Color and spatial structure in natural scenes. *Applied Optics*, 26(1):157–170, 1987.

[57] D. J. Field. Relations between the statistics of natural images and the response properties of cortical-cells. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 4(12):2379–2394, 1987.

[58] C. R. Carlson. Thresholds for perceived image sharpness. *Photographic Science and Engineering*, 22(2):69–71, 1978.

[59] M. S. Keil. "I Look in Your Eyes, Honey": internal face features induce spatial frequency preference for human face processing. *Plos Computational Biology*, 5(3), 2009.

[60] V. Goffaux and B. Rossion. Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology-Human Perception and Performance*, 33(4):995–1002, 2007.

[61] O. J. Hulme and S. Zeki. The sightless view: Neural correlates of occluded objects. *Cerebral Cortex*, 17(5):1197–1205, 2007.

[62] H. J. Lu and Z. L. Liu. When a never-seen but less-occluded image is better recognized: Evidence from old-new memory experiments. *Journal of Vision*, 8(7), 2008.

[63] D. A. Wilbraham, J. C. Christensen, A. M. Martinez, and J. T. Todd. Can low level image differences account for the ability of human observers to discriminate facial identity? *Journal of Vision*, 8(15):1–12, 2008.

[64] G. Wallis, U. E. Siebeck, K. Swann, V. Blanz, and H. H. Bulthoff. The prototype effect revisited: Evidence for an abstract feature model of face recognition. *Journal of Vision*, 8(3):1–15, 2008.

[65] Z. Solan and E. Ruppin. Similarity in perception: A window to brain organization. *Journal of Cognitive Neuroscience*, 13:18–30, 1999.

[66] R. M. Nosofsky and S. R. Zaki. A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual oldnew recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29:1194–1209, 2003.

[67] J. R. Solar and P. Navarrete. *Eigenspace-based Face Recognition.* Springer Engineering Series, 2001.

[68] A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.

[69] S. S. Rakover. Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology*, 93:1–30, 2002.

[70] N. Chater, J. B. Tenenbaum, and A. Yuille. Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7):287–291, 2006.

[71] S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7):682–687, 2002.

[72] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[73] Z. W. Tu and S. C. Zhu. Image segmentation by data-driven Markov Chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):657–673, 2002.

[74] Z. W. Tu, X. G. Chen, A. L. Yuille, and S. C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. *International Journal of Computer Vision*, 63(2):113–140, 2005.

[75] K. Friston. A theory of cortical responses. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360(1456):815–836, 2005.

[76] K. Friston. Learning and inference in the brain. *Neural Networks*, 16(9): 1325–1352, 2003.

[77] R. P. N. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[78] U. Grenander. *Elements of Pattern Theory.* Johns Hopkins University Press, 1996.

[79] S. Ullman. *High-Level Vision: Object Recognition and Visual Cognition.* MIT Press, 1996.

[80] D. Mumford. On the computational architecture of the neocortex .2. the role of corticocortical loops. *Biological Cybernetics*, 66(3):241–251, 1992.

[81] D. M. Mackay. Towards an information-flow model of human-behavior. *British Journal of Psychology*, 47(1):30–43, 1956.

[82] A. Yuille and D. Kersten. Vision as bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–308, 2006.

[83] T. R. Krynski and J. B. Tenenbaum. The role of causality in judgment under uncertainty. *Journal of Experimental Psychology*, 136(3):430–450, 2007.

[84] S.S.Intille and A.F.Bobick. Recognizing planned, multiperson action. *Computer Vision and Image Understanding*, 81:414445, 2001.

[85] R. Dahyot, P. Charbonnier, and F. Heitz. A Bayesian approach to object detection using probabilistic appearance-based models. *Pattern Analysis and Applications*, 7(3):317332, 2004.

[86] Y. Tong, W. Liao, and Q. Ji. Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1683–1699, 2007.

[87] F. Samaria and F. Fallside. Face identification and feature extraction using Hidden Markov Models. In *Image Processing: Theory and Applications*, pages 295–298. Elsevier, 1993.

[88] A. V. Nefian and M. H. Hayes. Hidden Markov models for face recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2721–2724, Seattle, USA, 1998.

[89] A. Nefian. Embedded Bayesian networks for face recognition. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 133–136, 2002.

[90] H. Othman and T. Aboulnasr. A separable low complexity 2D HMM with application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1229–1238, 2003.

[91] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[92] 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*, 19(7):711–720, 1997.

[93] M. S. Sarfraz and O. Hellwich. Probabilistic learning for fully automatic face recognition across pose. *Image and Vision Computing*, 28(5):744–753, 2010.

[94] M. S. Sarfraz and O. Hellwich. Learning probabilistic models for recognizing faces under pose variations. *Proceedings of the Image Mining Theory and Applications*, pages 122–132, 2008.

[95] J. L. Tu, Y. Fu, A. Ivanovic, T. S. Huang, and L. Fei-Fei. Variational transform invariant mixture of probabilistic pca. *IEEE Workshop on Applications of Computer Vision*, pages 21–26, 2008.

[96] Z. Lin and L. S. Davis. A pose-invariant descriptor for human detection and segmentation. *Computer Vision - Eccv 2008, Pt IV, Proceedings*, 5305: 423–436, 2008.

[97] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach toward pose-invariant face recognition. *2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Vols I-iii, Proceedings*, pages 954–959, 2003.

[98] M. Seshadrinathan and J. Ben-Arie. Pose invariant face detection. *Proceedings Ec-Vip-Mc 2003, Vols 1 and 2*, pages 405–410, 2003.

[99] L. Gu, S. Z. Li, and H. J. Zhang. Learning probabilistic distribution model for multi-view face detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 116–122, 2001.

[100] J. S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998.

[101] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[102] F. Matta and J. L. Dugelay. Person recognition using facial video information: A state of the art. *Journal of Visual Languages and Computing*, 20(3): 180–187, 2009.

[103] A. Doucet, N.D. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

[104] B. X. Li and R. Chellappa. A generic approach to simultaneous tracking and verification in video. *IEEE Transactions on Image Processing*, 11(5): 530–544, 2002.

[105] S. H. Zhou, V. Krueger, and R. Chellappa. Probabilistic recognition of human faces from video. *Computer Vision and Image Understanding*, 91 (1-2):214–245, 2003.

[106] B. Moghaddam. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780–788, 2002.

[107] S. H. K. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Transactions on Image Processing*, 13(11):1491–1506, 2004.

[108] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, 1998.

[109] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998.

[110] P. J. Huber. John W. Tukey's contributions to robust statistics. *Annals of statistics*, 30(6):1640–1648, 2002.

[111] K. C. Lee, J. Ho, M. H. Yang, and D. Kriegman. Visual tracking and recognition using probabilistic appearance manifolds. *Computer Vision and Image Understanding*, 99(3):303–331, 2005.

[112] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.

[113] A. Hadid and M. Pietikainen. Selecting models from videos for appearance-based face recognition, 2004.

[114] J. Kim, J. Choi, J. Yi, and M. Turk. Effective representation using ica for face recognition robust to local distortion and partial occlusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(12):1977–1981, 2005.

[115] J. Liu, S. Chen, D. Zhang, and X. Tan. An efficiient pseudoinverse linear discriminant analysis method for face recognition. *Intl. Journal of Pattern Recognition and Artificial Intelligence*, 21:1265–1278, 2007.

[116] J. Liu, S. C. Chen, X. Y. Tan, and D. Q. Zhang. Efficient pseudoinverse linear discriminant analysis and its nonlinear form for face recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(8): 1265–1278, 2007.

[117] G. Guo, S. Li, and K. Chan. Face recognition by support vector machines, 2000.

[118] M. H. Yang. Face recognition using kernel methods. *Advances in Neural Information Processing Systems 14, 2002, pp.*, 14:215–220, 2002.

[119] B. Schoelkopf, A. Smola, and K. R. Muller. Kernel Principal Component Analysis. In *In Artificial Neural Networks ICANN97*, 1997.

[120] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition, 2000.

[121] A. J. Howell and H. Buxton. Invariance in radial basis function neural networks in human face classification. *Neural Processing Letters*, 2:26–30, 1995.

[122] S. Lawrence, C. L. Giles, A.C.Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8:98–113, 1998.

[123] K. Hotta. Robust face recognition under partial occlusion based on support vector machine with local gaussian summation kernel. *Image and Vision Computing*, 26:14901498, 2008.

[124] A. M. Martinez and R. Benavente. The AR face database. CVC technical report no. 24. Technical report, 1998.

[125] A. M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.

[126] H. R. Kanan and K. Faez. Recognizing faces using adaptively weighted subgabor array from a single sample image per enrolled subject. *Image & Vision Computing*, 28(3):438–448, 2010.

[127] B.V.K.V. Kumar. Tutorial survey of composite filter designs for optical correlators. *Applied Optics*, 31(23):4773–4801, 1992.

[128] B.V.K.V. Kumar. *Correlation Pattern Recognition.* Cambridge University Press, Cambridge, UK, 2005.

[129] H. Laia, V. Ramanathana, and H. Wechsler. Reliable face recognition using adaptive and robust correlation filters. *Computer Vision and Image Understanding*, 111(3):329–350, 2008.

[130] W. C. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In *10th IEEE Intl. Conference on Computer Vision (ICCV2005)*, 2005.

[131] W. C. Zhang, S. G. G. Shan, and X. L. Chen. Local Gabor binary patterns based on Kullback-Leibler divergence for partially occluded face recognition. *IEEE Signal Processing Letters*, 14:875–878, 2007.

[132] H. J. Oh, K. M. Lee, and S. U. Lee. Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. *Image and Vision Computing*, 26:15151523, 2008.

[133] P.Penev and J. Atick. Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, 7:477–500, 1996.

[134] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Comput. Vis. Image Understanding 78 (2000)*, 78:99–118, 2000.

[135] S. Z. Li, X. W. Hou, and H. J. Zhang. Learning spatially localized, parts-based representation. *Computer Vision and Pattern Recognition*, 1:207–212, 2001.

[136] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y.Ma. Robust face recognition via sparse representation. *IEEE Trans. onPattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[137] R. Gross, I. Matthews, and S. Baker. Active appearance models with occlusion. *Image and Vision Computing*, 24:593–604, 2006.

[138] R. Gottumukkal and V. K. Asari. An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters*, 25(4):429–436, 2004.

[139] N. J. Nilsson. *Artificial Intelligence: A New Synthesis.* Morgan Kaufmann, 1998.

[140] S. Russel and P. Norvig. *Artificial Intelligence a modern approach.* Prentice Hall, 1995.

[141] A.P. Dawid. Conditional independence for statistical operations. *Annals of statistics*, 8:598–617, 1980.

[142] G. Davies, H. Ellis, and J. Shepherd. Cue saliency in faces as assessed by the photofit technique. *Perception*, 6:263–269, 1977.

[143] I. H. Fraser, G. L. Craig, and D. M. Parker. Reaction time measures of feature saliency in schematic faces. *Perception*, 19(5):661–673, 1990.

[144] H. D. Ellis, J. W. Shepherd, and G. M. Davies. Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8(4):431–439, 1979.

[145] A. W. Young, D. C. Hay, B. M. Flude K.H. McWeeny, and A. W. Ellis. Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14:737–746, 1985.

[146] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Prentice Hall, 2002.

[147] R. E. Neapolitan. *Learning Bayesian Networks.* Prentice Hall, 2003.

[148] Papoulis. *Probability, random variables, and Stochastic Processes.* McGrawHill, New York, 2002.

[149] S. Haykin. *Neural Networks: A comprehensive foundation.* Prentice Hall, New Jersey, 1999.

[150] G. W. Cottrell and J. Metcalfe. Face, gender and emotion recognition using holons, 1991.

[151] 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*, 4(3):519–524, 1987.

[152] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[153] A. Pentland, B. Moghaddam, and T. Starner. Viewbased and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition*, pages 84–91, Seattle, US, 1994. IEEE Computer Society.

[154] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2001.

[155] J. Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47:90–100, 2003.

[156] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[157] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

[158] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, 15:252–260, 2006.

[159] K. Murphy. Software for graphical models: a review. Technical report, International Society for Bayesian Analysis (ISBA) Bulletin, December 2007.

[160] A. K. Jain and B. Chandrasekaran. *Dimensionality and sample size considerations in pattern recognition practice - Handbook of Statistics*, volume 2. Elsevier, 1982.

[161] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1):4–3, 2000.

[162] P.J.Phillips, H.Wechsler, and P.Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing Journal*, 16:295–306, 1998.

[163] M. S. Nixon and J. N. Carter. Automatic recognition by gait. *IEEE special issue on Biometrics: Algorithms & Applications*, 94:2013–2024, 2006.

[164] Z. Liu and S. Sarkar. Effect of silhouette quality on hard problems in gait recognition. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 35:170–183, 2005.

[165] C. Bauckhage, J. K. Tsotsos, and F. E. Bunn. Automatic detection of abnormal gait. *Image and Vision Computing*, 27:108115, 2006.

[166] E. Muybridge. *The Human Figure in Motion.* Dover, New York, 1901.

[167] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14:201–211, 1973.

[168] D. Gafurov and E. Snekkenes. Gait recognition using wearable motion recording sensors. *Eurasip Journal on Advances in Signal Processing*, 2009.

[169] I. Bouchrika and M. S. Nixon. Gait recognition by dynamic cues. *19th International Conference on Pattern Recognition, Vols 1-6*, pages 189–192, 2008.

[170] H. Lakany. Extracting a diagnostic gait signature. *Pattern Recognition*, 41 (5):1627–1637, 2008.

[171] M. S. Nixon and J. N. Carter. Automatic recognition by gait. *Proceedings of the Ieee*, 94(11):2013–2024, 2006.

[172] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90(1):1–41, 2003.

[173] A. Veeraraghavan, A. Srivastava, and K. R. Chowdhury. Rate-invariant recognition of humans and their activities. *IEEE Transactions on Image Processing*, 18(6):1326–1339, 2009.

[174] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. *IEEE Trans. on Systems, Man and Cybernetics - Part C: Applications and Reviews VOL. 38, NO. 2, MARCH 2008*, 38:145–155, 2008.

[175] L. T. Kozlowski and J. E. Cutting. Recognizing sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21(6):575–580, 1977.

[176] W. H. Dittrich. Action categories and the perception of biological motion. *Perception*, 22(1):15–22, 1993.

[177] G. P. Bingham, R. C. Schmidt, and L. D. Rosenblum. Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology-Human Perception and Performance*, 21(6):1473–1493, 1995.

[178] G. L. Pellecchia and G. E. Garrett. Assessing lumbar stabilization from point-light and normal video displays of manual lifting. *Perceptual and Motor Skills*, 85(3):931–937, 1997.

[179] S. V. Stevenage, M. S. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, 13(6):513–526, 1999.

[180] J. E. Boyd. Synchronization of oscillations for machine perception of gaits. *Computer Vision and Image Understanding*, 96(1):35–59, 2004.

[181] B.I. Bertenthal and J. Pinto. A dynamic systems approach to development applications. In L.B.Smith and E. Thelen, editors, *Complementary processes in the perception and production of human movements*, pages 209–239. MIT Press, Cambridge, 1993.

[182] S. H. Collins, M. Wisse, and A. Ruina. A three-dimensional passive-dynamic walking robot with two legs and knees. *International Journal of Robotics Research*, 20(7):607–615, 2001.

[183] M. J. Coleman and A. Ruina. An uncontrolled walking toy that cannot stand still. *Physical Review Letters*, 80(16):3658–3661, 1998.

[184] M. Garcia, A. Chatterjee, A. Ruina, and M. Coleman. The simplest walking model: Stability, complexity, and scaling. *Journal of Biomechanical Engineering-Transactions of the Asme*, 120(2):281–288, 1998.

[185] T. McGeer. Passive dynamic walking. *International Journal of Robotics Research*, 9(2):62–82, 1990.

[186] D. K. Wagg and M. S. Nixon. On automated model-based extraction and analysis of gait. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings*, pages 11–16, 2004.

[187] Z. Zhou, A. P. Bennett, and R. I. Damper. A bayesian framework for extracting human gait using strong prior knowledge. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(11):1738–1752, 2006.

[188] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. R. Chowdhury, K. Volker, and R. Chellappa. Identification of humans using gait. *IEEE Trans. on Image Processing*, 13:1163–1173, 2004.

[189] L. Wang and T. Tan. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25: 1505– 1518, 2003.

[190] I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1998.

[191] J. Kent. *New directions in shape analysis*. Wiley, Chichester, 1992.

[192] L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification, 2002.

[193] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. *Fifth Ieee International Conference On Automatic Face And Gesture Recognition, Proceedings*, pages 372–377, 2002.

[194] CASIA. Gait database offered by Chinese Academy of Sciences, http://www.sinobiometrics.com, 2006.

[195] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos. Gait recognition using linear time normalization. *Pattern Recognition*, 39:969–979, 2006.

[196] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18-39(1), 2008.

[197] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer, 2006.

[198] A. Veeraraghavan, A. R. Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis, June 2004.

[199] Y. Yang and M. Levine. The background primal sketch: An approach for tracking moving objects. *Machine Vision and Applications*, 5:17–34, 1992.

[200] C. H. Chen, J. M. Liang, H. Zhao, H. H. Hu, and J. Tian. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 30(11):977–984, 2009.

[201] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon. A floor sensor system for gait recognition. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies, Proceedings*, pages 171–176, 2005.

[202] J. Jenkins and C. Ellis. Using ground reaction forces from gait analysis: Body mass as a weak biometric. *Pervasive Computing, Proceedings*, 4480: 251–267, 2007.

[203] H. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S. M. Makela. Identifying people from gait pattern with accelerometers. *Biometric Technology for Human Identification II*, 5779:7–14, 2005.

[204] B. F. Huang, M. Chen, P. F. Huang, and Y. S. Xu. Gait modeling for human identification. *Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Vols 1-10*, pages 4833–4838, 2007.

[205] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng. A wearable acceleration sensor system for gait recognition. *2nd IEEE Conference on Industrial Electronics and Applications*, 1:2654–2659, 2007.

[206] M. Sekine, Y. Abe, M. Sekimoto, Y. Higashi, T. Fujimoto, T. Tamura, and Y. Fukui. Assessment of gait parameter in hemiplegic patients by accelerometry. *Proceedings of the 22nd Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-4*, 22:1879–1882, 2000.

[207] D. Alvarez, R. C. Gonzalez, A. Lopez, and J. C. Alvarez. Comparison of step length estimators from weareable accelerometer devices. *2006 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-15*, pages 2135–2138, 2006.

[208] H. Moon and P. J. Phillips. Computational and performance aspects of PCA-based face recognition algorithms. *Perception*, 30:303321, 2001.

[209] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang. Discriminant analysis with tensor representation, 2005.

[210] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006.

[211] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.J.Zhang. Human gait recognition with matrix representation. *IEEE Transactions on circuits and systems for video technology*, 16(7):896–903, 2006.

[212] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, 2006.

[213] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann Publishers, 1997.

[214] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 63:107–136, 2006.

[215] P. Domingos, D. Lowd, and R. Barchman. *Markov Logic: An Interface Layer for Artificial Intelligence.* Morgan and Claypool, 2009.

[216] N. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[217] F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge.* MIT Press, Cambridge, 1990.

[218] J. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311350, 1990.

[219] J. S. Wellman, M. Breese, and R. P. Goldman. From knowledge bases to decision models. *Knowledge Engineering Review*, 7:3553, 1992.

[220] S. Kok and P. Domingos. Learning the structure of Markov logic networks. pages 441–448, Germany, 2005. ACM Press.

[221] J. Neville and D. Jensen. Dependency networks for relational data. In *In Proceedings of the Fourth IEEE International Conference on Data Mining,*, pages 170–177, Brighton, UK, 2004. IEEE Computer Society Press.

[222] N. Friedman, L. Getoor, and D. Koller. Learning probabilistic relational models. In *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, Sweden, 1999. Morgan Kaufmann.

[223] S. Muggleton. Stochastic logic programs. In L. De Raedt, editor, *Advances in Inductive Logic Programming*, pages 254–264, 1996.

[224] S. D. Tran and L. S. Davis. Event modeling and recognition using Markov logic networks. *Computer Vision - ECCV 2008, Pt II, Proceedings*, 5303: 610–623, 2008.

[225] V. D. Shet, D. Harwood, and L. S. Davis. Multivalued default logic for identity maintenance in visual surveillance. *Computer Vision - ECCV 2006, Pt 4, Proceedings*, 3954:119–132, 2006.

[226] N. A. Rota and M. Thonnat. Activity recognition from video sequences using declarative models. *ECAI 2000: 14th European Conference on Artificial Intelligence, Proceedings*, 54:673–677, 2000.

[227] F. Wu and D. Weld. Automatically refining the wikipedia infobox ontology. In *Proceedings of the 2008 International Conference on the World Wide Web*, pages 635–644, China, 2008.

[228] D. Lowd and P. Domingos. Efficient weight learning for Markov logic networks. In *European Conference of Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD-2007)*, pages 200–211, 2007.

[229] M. R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, San Mateo, CA, 1987.

[230] H. X. Jia, C. Moore, and B. Selman. From spin glasses to hard satisfiable formulas. In H. H. Hoos and D. G. Mitchell, editors, *Theory and Applications of Satisfiability Testing*, volume 3542 of *Lecture Notes in Computer Science*, pages 199–210. Springer-Verlag Berlin, Berlin, 2005.

[231] M. Fitting. *First-Order Logic and Automated Theorem Proving.* Springer-Verlag, New York, 1990.

[232] T. G. Dietterich, P. Domingos, L. Getoor, S. Muggleton, and P. Tadepalli. Structured machine learning: the next ten years. *Machine Learning*, 73(1): 3–23, 2008.