Dynamic Descriptors in Human Gait Recognition

by

Tahir Amin

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2013 by Tahir Amin
Abstract

Dynamic Descriptors in Human Gait Recognition

Tahir Amin

Doctor of Philosophy

Graduate Department of Electrical and Computer Engineering

University of Toronto

2013

Feature extraction is the most critical step in any human gait recognition system. Although gait is a dynamic process yet the static body parameters also play an important role in characterizing human gait. A few studies were performed in the past to assess the comparative relevance of static and dynamic gait features. There is, however, a lack of work in comparative performance analysis of dynamic gait features from different parts of the silhouettes in an appearance based setup. This dissertation presents a comparative study of dynamic features extracted from legs, arms and shoulders for gait recognition. Our study partially supports the general notion of leg motion being the most important determining factor in gait recognition. But it is also observed that features extracted from upper arm and shoulder area become more significant in some databases. The usefulness of the study hinges on the fact that lower parts of the leg are generally more noisy due to a variety of variations such as walking surface, occlusion and shadows. Dynamic features extracted from the upper part of the silhouettes possess significantly higher discriminatory power in such situations. In other situations these features can play a complementary role in the gait recognition process.

We also propose two new feature extraction methods for gait recognition. The new methods use silhouette area signals which are easy and simple to extract. A significant performance increase is achieved by using the new features over the benchmark method and recognition results compare well to the other current techniques. The simplicity and
compactness of the proposed gait features is their major advantage because it entails low computational overhead.
Dedication

To my family
Acknowledgements

First and foremost, I would like to extend my sincere thanks to my supervisor Prof. Dimitrios Hatzinakos for his encouragement, guidance, and continuous support throughout my research work and writing of this manuscript. This work would have been impossible without his feedback, patience and kindness. I consider it an honour to work with him.

I wish to express my gratitude to my thesis advisory committee members, Prof. Konstantinos N. Plataniotis and Prof. Parham Aarabi, for providing me valuable feedback and advice during my research work and writing of this manuscript. I would also like to thank Prof. Glenn Gulak for his insightful input and comments on my thesis. My sincere gratitude is due to Prof. Svetlana Yanushkevic for sparing her valuable time to serve as external examiner for my final oral examination. My thanks are due to the Department of Electrical and Computer Engineering, University of Toronto for providing research and computing resources.

I would like to acknowledge my supervisor’s funding resources National Sciences and Engineering Research Council of Canada (NSERC) for financial support to this research work. I would also like to acknowledge the Ontario Graduate Scholarship (OGS) program for providing me financial support during the early years of my research work.

I am thankful to my colleagues and members of the Multimedia Laboratory for creating a friendly and congenial environment in the Lab. It was a pleasure to work with Dr. Yongjin Wang, Dr. Francis Bui, Dr. Foteini Agrafioti and other graduate students.

Last but not least, I owe my deepest gratitude to my family, my parents, siblings, beloved wife and son Daniel for their unconditional love, encouragement, understanding and support. Their love has kept me going through some of the most difficult and challenging times of my life.
Contents

List of Tables x

List of Figures xii

List of Abbreviations xv

1 Introduction 1

1.1 Motivation and Applications . . . . . . . . . . . . . . . . . . . . . . . . 3
1.2 Biometric Systems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
1.3 Gait Biometric . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
1.4 Performance Metrics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
1.5 Body Dynamics and Gait Recognition . . . . . . . . . . . . . . . . . . . 16
1.6 Contributions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
1.7 Thesis Organization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

2 Literature Review 21

2.1 Model Based Gait Recognition . . . . . . . . . . . . . . . . . . . . . . . . 23
2.2 Appearance Based Gait recognition . . . . . . . . . . . . . . . . . . . . . 29
2.3 Miscellaneous Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
2.4 Dynamic versus Static Features . . . . . . . . . . . . . . . . . . . . . . . 41
2.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
List of Tables

2.1 Model based gait recognition research ........................................ 30
2.2 Appearance based gait recognition techniques ............................... 39

3.1 Summary of gait databases ......................................................... 52
3.2 GTech probe sets ..................................................................... 53
3.3 Number of sequences for each possible combination .......................... 56
3.4 Probe set for each of the gait challenge experiments ........................... 57
3.5 Probe set for gait challenge experiments A–G ................................. 58
3.6 Comparison of features at rank 1 and rank 5 for GC database .......... 60
3.7 Comparison of features at rank 1 and rank 5 for GTech database ........ 62

4.1 Recognition results for probe 1 ..................................................... 73
4.2 Recognition results for probe 3 ..................................................... 75
4.3 Recognition results for probe 4 ..................................................... 75
4.4 Recognition results for probe 5 ..................................................... 78

5.1 Comparison of db1, db2 and db4 based features at rank 1 ............... 84
5.2 Comparison of db1, db2 and db4 based features at rank 5 ............... 85
5.3 Performance comparison for probe 1 ........................................... 86
5.4 Performance comparison for probe 2 ........................................... 87
5.5 Performance comparison for probe 3 ........................................... 87
5.6 Performance comparison for probe 4 ........................................... 88
5.7 Performance comparison for probe 5
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Block diagram of a gait verification system</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>Block diagram of a gait identification system</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>Biometric sensors</td>
<td>9</td>
</tr>
<tr>
<td>1.4</td>
<td>Biometric System Errors</td>
<td>14</td>
</tr>
<tr>
<td>1.5</td>
<td>ROC curve showing relationship between FRR and FAR</td>
<td>15</td>
</tr>
<tr>
<td>1.6</td>
<td>A Typical CMC</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>Contours of walking and running subject with point configuration</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Human gait cycle</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>Area masks</td>
<td>45</td>
</tr>
<tr>
<td>3.3</td>
<td>Area signals for a silhouette sequence</td>
<td>46</td>
</tr>
<tr>
<td>3.4</td>
<td>Area signals for another silhouette sequence</td>
<td>47</td>
</tr>
<tr>
<td>3.5</td>
<td>Noisy area signals</td>
<td>48</td>
</tr>
<tr>
<td>3.6</td>
<td>Input signal $a_1$ and its IMFs</td>
<td>49</td>
</tr>
<tr>
<td>3.7</td>
<td>Input signal $a_2$ and its IMFs</td>
<td>50</td>
</tr>
<tr>
<td>3.8</td>
<td>Area signals after noise removal</td>
<td>51</td>
</tr>
<tr>
<td>3.9</td>
<td>Example video frames from GTech database (a) Indoor angle (b) Indoor side near angle (c) Indoor side far angle (d) Outdoor angle</td>
<td>54</td>
</tr>
<tr>
<td>3.10</td>
<td>Example silhouettes from GTech database</td>
<td>55</td>
</tr>
</tbody>
</table>
3.11 Example video frames from GC database (a,b) Concrete surface (b,c)
Grass surface ................................................................. 55
3.12 Example silhouettes from GC database ................................ 56
3.13 CMC of six area based features for GC database ................. 61
3.14 CMC of six area based features for GTech database ............. 63
4.1 Silhouettes containing outliers ....................................... 65
4.2 Silhouettes after median filtering .................................... 65
4.3 Morphological closing on a silhouette ............................... 66
4.4 Silhouette area signal .................................................. 68
4.5 Autocorrelation of silhouette area signal ............................ 69
4.6 CMC curve for probe 1 .................................................. 74
4.7 CMC curve for probe 3 .................................................. 76
4.8 CMC for probe 4 ......................................................... 77
4.9 CMC for probe 5 ......................................................... 78
5.1 Sampling grid for time-scale plane ................................... 80
5.2 STFT coverage of Time-Frequency plane ............................ 80
5.3 CMC curves for level 1 to level 5 decomposition .................... 86
List of Abbreviations

CASIA  Chinese Academy of Sciences
CCD  Centroid Contour Distance
CF  Correlation Features
CMC  Cumulative Match Curve
CMS  Cumulative Match Score
CMU  Carnegie Mellon University
CWT  Continuous Wavelet Transform
db  Daubechies
DCT  Discrete Cosine Transform
DF  Distance Features
DFT  Discrete Fourier Transform
DOF  Degree Of Freedom
DWT  Discrete Wavelet Transform
EM  Expectation Maximization
EMD  Empirical Mode Decomposition
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERR</td>
<td>Equal Error Rate</td>
</tr>
<tr>
<td>FA</td>
<td>False Accept</td>
</tr>
<tr>
<td>FAR</td>
<td>False Accept Rate</td>
</tr>
<tr>
<td>FD</td>
<td>Fourier Descriptor</td>
</tr>
<tr>
<td>FR</td>
<td>False Reject</td>
</tr>
<tr>
<td>FRR</td>
<td>False Reject Rate</td>
</tr>
<tr>
<td>GC</td>
<td>Gait Challenge</td>
</tr>
<tr>
<td>GTech</td>
<td>Georgia Tech</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>IMF</td>
<td>Intrinsic Mode Function</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LTN</td>
<td>Linear Time Normalization</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum a Posteriori</td>
</tr>
<tr>
<td>MLD</td>
<td>Moving Light Display</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristics</td>
</tr>
<tr>
<td>SOTON</td>
<td>University of Southampton</td>
</tr>
<tr>
<td>STFT</td>
<td>Short Time Fourier Transform</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

There has been a phenomenal growth in the digital video data during the last decade. This is a result of cheaper digital devices to acquire digital image and video footage and higher storage capacities available with the Personal Computers (PC). Moreover, the processing power of the PCs is also increasing at exponential rates according to the Moore’s Law [1]. This increased computational power has made it possible to process and analyze huge data sets that was unthinkable in the recent past. Video is a rich source of information with applications to a variety of areas. It is inconceivable to imagine the modern entertainment industry without video. Apart from entertainment, video is used in a variety of other areas such as surveillance, security, traffic monitoring and medical diagnostics. A large amount of video data is now available in digital format or can easily be converted into digital format. Therefore it can be manipulated and analyzed using computers.

The recognition of people by their physiological or behavioral characteristics is called biometrics. There are several biometrics that are being used for personal identification such as fingerprints, DNA, face, retinal scan, iris, voice, foot, hand geometry and gait. Any physical or behavioural trait may be used as a biometric as long as it fulfills the following four requirements [2].
Chapter 1. Introduction

- **Universality:** It should be universal meaning all individuals should have the characteristic

- **Distinctiveness:** It should be sufficiently distinctive among different individuals

- **Permanence:** The characteristic should remain invariant over a period of time

- **Collectability:** It should be quantitatively measurable.

In addition to the above mentioned requirements, the practical biometric systems also entail the issues of performance, acceptability, circumvention, privacy and data security. A practical biometric system should fulfil certain performance requirements in terms of its accuracy, speed, reliability and resource consumption. It should also be acceptable for the population and highly robust to the fraudulent methods and attacks. The security of the biometric data is also a fundamental requirement because of privacy issues and the richness of information associated with the biometric data.

The use of biometrics in personal identification is not new and it has been used in criminology for a long time. Even before the advent of modern technology for acquisition and archiving of fingerprint databases, biometrics such as the artistic drawings of facial features, color of eyes and footprints were routinely used in personal identification especially in criminology. Fingerprint databases have widely been used by the law enforcement agencies from the early 19th century. The invention of modern computers and digital technology have, however, transformed the way information is stored and processed. Digital technology has enabled us to store and process the biometric data automatically without human intervention or with minimal human input. The renewed focus on security in the past few years has brought the biometrics research into limelight. Recent developments in the biometric research has brought face, iris and fingerprint recognition from research labs to daily life. Biometric recognition systems are being installed as access control systems for granting access to offices, buildings and even laptop computers.
1.1 Motivation and Applications

The traditional methods of user identification such as passwords, user IDs, magnetic strip cards and PIN codes suffer from many drawbacks. The main disadvantage of these methods is that they test the validity of the password, PIN or magnetic card and not the actual user. The reason is that they can easily be shared with illegitimate users. The genuine users can share their passwords, PIN codes and ID cards with their friends and family to give them access to the system. Passwords and PIN codes can be stolen and lost magnetic cards can be used to get illegitimate access. The strength of PIN codes and passwords greatly depends on the users themselves because they are allowed to choose their own passwords and PIN codes to automate the enrollment process. The majority of users choose a password that can be guessed easily if you have a little information about the user. It has been observed that people choose obvious words as passwords such as their birth dates, birth dates of their children and pet’s names.

The result of a security breach can be catastrophic because it is not limited to just computers. It can give access to a lot of other resources such as bank accounts, personal tax records, medical records and national security data. Many techniques are used to make the usage of passwords and PIN codes more secure. Their storage and transmission is encrypted so they cannot be deciphered that easily. However, the trojans and viruses that appear as legitimate attachments in emails are automatically installed on computers and start logging keystrokes and sending all this information to the creators. This type of hacking attacks can make all the encryption schemes useless. Although conventional access control systems are vulnerable in many ways yet there are some advantages of using them. The passwords and PIN codes can be changed easily if they become compromised. Similarly the ID cards can be cancelled and new cards can issued. These system generally need low resources and are cheaper and easier to maintain. As we will see in later discussion that this is not the case with biometrics.

The problems with traditional access control methods paved the way to design better
and more secure systems. In the age of Internet, social media and E-services, a security
breach can expose almost all aspects of personal and financial history and information.
The issues and shortfalls of traditional access control methods and the renewed focus
on security in the recent past has led to the development of biometric technology. The
major advantage of biometrics is that the ID is part of users’ physiology or behaviour
and is strictly attached to them. Hence it cannot be moved, exchanged or stolen in a way
that passwords and PIN codes are. There are no passwords or PIN codes to remember
and a need to carry security cards in certain setups. Biometrics are also more secure to
brute force attacks unlike passwords which are prone to even dictionary attacks.

Biometric technology comes with a lot of benefits and has solved several issues faced
by traditional access control systems but it has its own drawbacks. One of its main
drawback is that we only have a limited number of biometrics such as two hands, one
face, two irises and two retinas. The re-issuance of IDs is not a straight forward affair
like passwords and PIN codes. Biometrics systems process a lot more data and perform
complex computations compared to traditional systems. They need more resources and
are generally more expensive. The privacy of the users is also a major concern in biometric
systems. The biometric data in itself carries a lot of private information unlike PIN codes
or passwords that do not contain any personal information by itself and are merely a mean
to access it. Biometrics can reveal a lot of sensitive information about the individuals
such as race, medical history and risk factors for certain diseases. This can have drastic
effects on the individual’s well being and access to services if got into wrong hands. The
insurance companies for example can deny coverage to some individuals based on this
information. The questions like who can access biometric data? how is it stored and used?
are still open questions. There are also cases where certain individuals will not possess
one kind of biometric due to injury, heredity or type of work. People working in some
chemical industries may lose fingerprints or the quality of fingerprints is not sufficient for
enrollment. In these cases, it will be required to deploy technologies capable of working
with multiple biometrics only to increase the cost of installation and maintenance.

The major application of biometrics is in the access control systems as discussed in the preceding paragraphs. Applications such as building access control, access control to electronic resources like PCs, laptops, bank machines, personal security locks are all benefiting from biometric technology. Currently, face, iris and fingerprint are the most popular and reliable choice for this kind of applications. Digital cameras with face detection and recognition provide better picture quality and automatic tagging of digital photos and video.

Gait analysis and recognition from digital video data has many applications. For example, digital video footage from subway stations, airports or other public places may be processed to find event exceptions in real time. Situations like a running person in a crowded subway station are uncommon and can be detected by extracting the motion/activity descriptors of the human objects. This can serve as an early warning for the law enforcement agencies or trigger an alarm to alert the security personnel. Similarly, studies have shown that humans have unique gait signatures which are not only different from the animals but they are also differentiable within the human race. It has been observed that the human motion contains strong periodicities while the animal motion does not show such characteristics. This property of gait can be used for object extraction and classification from digital videos for indexing and retrieval applications. Video data can be summarized and segmented using gait based descriptors for the intelligent browsing of the videos. Human gait analysis can be implemented to serve as identity verification for access control or criminology. Another promising application of such analysis is the medical diagnostics of diseases that affect the voluntary muscle activity such as walking. For example, Parkinson’s disease affects nerve cells (neurons) in the part of the brain controlling muscle movement. People with Parkinson’s often experience trembling, muscle rigidity, difficulty in walking, and problems with balance and coordination. Early detection of walking disorders by the motion analysis can be
very helpful for the treatment of such diseases.

1.2 Biometric Systems

A biometric system is essentially a pattern recognition system which recognizes users by comparing their specific anatomical or behavioral characteristic with stored templates. The users must be enrolled in the system so that his biometric template or reference can be captured. This template is securely stored in a central database or a smart card issued to the user. The template is used for matching when an individual needs to be identified. Depending on the context, a biometric system can operate either in verification (authentication) mode or identification mode.

- **In the verification** (Am I who I claim I am?) mode, the system performs 1:1 match between the stored template and the probe template to confirm or deny the claimed identity.

- **In identification** (Who am I?) mode, the system has to perform 1:N matches between probe template and all the N templates stored in the database.

Identification is a more challenging problem because it involves 1:N matching compared to 1:1 matching for verification. Figures 1.1 and 1.2 show the block diagrams of a typical verification and identification system respectively.

A typical biometric system consists of five fundamental modules.

1. **Sensor**: This module collects the biometric data for further processing by the system. It essentially consists of sensors for data collection. In monocular gait recognition systems, it consists of one video camera. More than one video camera is needed for 3D data collection which is used for stereo scene generation and gait recognition. Still cameras are used for collection of image data for face recognition. The still cameras used for iris recognition capture high resolution images necessary
for extraction of quality signatures for iris recognition. Fingerprint readers are used for collection of fingerprints in the fingerprint recognition systems. In some applications the pre-collected data may also be used for biometric recognition and the biometric system may work without this module. This scenario is common in criminal investigations where the ubiquitous personal videos and images found on social networking sites such as Facebook and YouTube are being used to find clues to the activities of criminals as well as that of victims. The existing security videos captured for surveillance purposes at the subway stations, airports and other public and private locations may also be processed by a gait or face recognition system. Figure 1.3 shows the images of sensors used to collect ekg, face, fingerprint, gait, iris and palm data.

2. **Pre-Processor:** In some cases, biometric data collected by the sensor module is not suitable for feature extractor module. For example, the majority of feature extraction algorithms for gait recognition work with silhouette sequences rather than
raw video sequences captured by gait cameras. The pre-processor module extracts the silhouettes from video sequences in this kind of gait recognition systems. In addition to the data format conversion, there may be additional pre-processing performed by this module such as noise removal, quality check and improvement. The silhouettes obtained from segmentation algorithms are noisy due to the shadows, occlusion of body parts, variation in lighting and other environmental conditions. Different filtering techniques are applied to remove the noise from the silhouettes. Morphological operations may be applied for thinning and filling the holes in the silhouettes during the pre-processing. There might be situations where the quality of some silhouettes is so low that it is impossible to extract reliable features from them. These silhouettes can be excluded from the feature extraction process and only those data is used that provide better quality and robust features. The pre-processor module may perform some kind of quality check on the silhouettes and decides whether to include or exclude it for feature extraction.

3. **Feature Extractor:** The biometric data such as raw gait video sequences is huge and its storage and processing in this form will require very high computational
resources. The majority of applications such as access control system require an immediate response. Therefore the storage and comparison of video sequences is neither feasible nor practical. There may also be hidden characteristics in the data that are not apparent in its raw format. The transformation of the data in a different domain may expose those characteristics. Moreover, there is usually a lot of redundant and irrelevant information in the data that affect the system efficiency negatively. Feature extraction is performed on the data to acquire a reliable, compact and efficient representation of the biometric. After pre-processing of the captured biometric data, discriminatory features are extracted from it by the feature extractor module. A biometric template is generated from these extracted features and it serves as a signature or ID of the individual. This is the most critical module in any biometric recognition system. The quality and size of features will have the single largest effect on the performance and speed of the biometric
system. A compact and robust feature set is desirable to obtain efficient and reliable performance from the system. For example, in gait recognition systems, joint angle trajectories, areas of different body segments, distances between body segments during walking process and a variety of other characteristics are used as features. This module works in the offline mode during the enrollment process and in the online mode during the recognition stage.

4. Pattern Matcher: This module performs the pattern matching between the stored biometric templates and the features obtained during the recognition phase. When a user walks in front of a gait camera, the resulting video data is processed by the pre-processor and feature extractor modules to obtain the same set of features that were stored during the enrollment. This feature set is then compared with the stored templates and a match score is usually obtained by this module. Now depending on the system accuracy settings, a decision is made to either accept or reject the claimed identity in the verification scenario. The system gives the established identity in an identification setup. Some applications may need a human input in the decision process. The best matches are displayed by the system based on the match scores and a human expert makes the final decision. In other applications such as criminal investigations, the person may not be present in the database and no matches will be found. The process is similar in case of other biometrics such as fingerprint, face and iris. The difference lies only in the acquisition of biometric data.

5. Feature Database: Biometric templates of all the users must be stored by enrollment process before they can use the system. These templates may be stored in a central database or on the smart cards carried by the users. It has been observed that the biometric data of the same person varies with the passage of time and other conditions. Therefore, it is usually desirable to store more than one template of the
same person in the database or on the smart card. The security of the biometric database is also of much concern these days and different encryption technologies are employed to store the templates securely.

1.3 Gait Biometric

Gait is a behavioral biometric which can be perceived from a distance. It can be acquired without personal contact and cooperation. Iris and face biometrics have similar advantages but they need high resolution images and the frontal view. However, it is possible to extract gait patterns from low resolution images. Human gait can vary over long durations due to the change in body weight, injuries and disease. But studies have indicated that it still possesses sufficient discriminatory power for personal recognition [4]. Gait is a complex function of skeletal structure, muscular activity, body weight, limb lengths and bone structures etc. This complexity of gait renders it difficult to imitate and hide if not impossible. The performance of gait biometric for verification and recognition lags behind some other biometrics such as DNA, iris and face. This is due to the fact that individuals’ gait shows larger variations because it is a behavioural characteristic. Therefore it is more suitable for low security applications. Socially the collection of gait data is highly acceptable. The widespread usage of video surveillance cameras at public and private premises indicates its acceptability. Video recordings from the surveillance cameras are of such a low resolution that face recognition cannot be performed. In these circumstances, gait can aid in personal recognition and can provide valuable information in many situations such as crime scene investigations and customers’ visiting patterns.

1.4 Performance Metrics

The biometric data of the same individual collected at different occasions varies due to a number of factors and no two samples are exactly the same. This is due to the
very nature of the biometric data as well as the conditions under which it is collected. Physiological and behavioural changes in the individuals can affect different biometrics to varied degrees. Aging affects most of the biometrics in some way. The facial appearances, retina, gait, finger prints and other biometrics show variations with age. The noise in the acquisition process and environmental conditions also affect the collected data. For example, video gait data shows variations with the change of walking surface. The individual’s walking pattern may be different on the hard surface vs on the soft surface. Walking on grass will introduce noise such as occlusion of feet. Shadows during data acquisitions will also introduce noise. Similarly variation in lighting conditions will effect all image based biometrics. Injury will also affect biometrics and in some cases can render some biometrics completely unusable.

Due to the variations in the biometric data as discussed in the preceding paragraph, the pattern matching process cannot find an exact match. Hence a similarity measure is defined which determines how closely a collected feature matches with the stored template. A threshold is also defined and the decision is based on whether the similarity value falls within or outside this threshold. The value of the threshold is tweaked according to the application and purpose of the biometric system. If the threshold allows very small variations, even the identity of the genuine users may be rejected by the system. On the other hand if the threshold allows larger variations between the template and the features extracted during the recognition phase, the system may verify non enrolled users. A very careful investigation of the behaviour of the system is performed to set the optimal threshold to achieve a balance between performance, error and usability.

The identity of the individual is represented by his/her biometric sample obtained and stored during the enrollment process. This sample is most commonly a compact feature template extracted by the feature extractor module from the acquired biometric data. Let us say that we have $N$ users of the system that have all been enrolled. The template database is commonly called the gallery set in gait recognition community. Each template
Chapter 1. Introduction

in the gallery set, $g_k, k = 1, 2, \cdots, N$, represents the identity of the person $I_k$. During
the authentication stage, probe feature vector $p$ is extracted from the biometric data of
the person $I_p$ claiming the identity $I_v$. A decision to accept or reject the claim is based
on the similarity measure ($S$) between the probe feature vector ($p$) and the template
feature vector ($g_v$) of the claimed identity $I_v$ in the gallery database. The process can
formally be posed as a hypothesis testing problem as follows [2, 3].

$$H_0 : S(p, g_v) \geq T, \text{ the claimed identity is accepted}$$

$$H_1 : S(p, g_v) < T, \text{ the claimed identity is rejected} \quad (1.1)$$

where $T$ is the threshold.

A verification biometric system operates in a binary decision mode i.e. the claimed
identity is either accepted or rejected. Therefore it can make two types of errors.

- **Type I** error occurs when a genuine identity claim is rejected by the system. In
terms of Equation 1.1, $H_1$ is accepted when $H_0$ is true. This error is also called
False Reject (FR) error.

- **Type II** error occurs when a false identity claim is accepted by the system. In
terms of Equation 1.1, $H_0$ is accepted when $H_1$ is true. This error is also called
False Accept (FA) error.

**Type I and Type II** errors are measured by False Reject Rate (FRR) and False Accept
Rate (FAR) respectively. These two error rates characterize any recognition system and
are closely tied to the threshold value, $T$. FRR and FAR are defined below.

**FRR** is the probability of deciding $H_1$ when $H_0$ is true. It is calculated as follows:

$$FRR = \frac{1}{N} \sum_{i=1}^{N} FRR(n) \quad (1.2)$$

$$FRR(n) = \frac{\text{False rejected transactions for person } n}{\text{Total transactions of person } n} \quad (1.3)$$
Where $FRR(n)$ represents the FRR value for person $n$ and FRR is the average over all users in the database. FRR depends not only on the biometric system but also on the user and how the user interacts with the system. This is a statistical measure and its accuracy depends on the number of measurements. Now in order to avoid the calculation being biased by a person, first the personal FRR is calculated for each person. Then the average is obtained as per Equation 1.3.

$FAR$ is the probability of deciding $H_0$ when $H_1$ is true. It is calculated as follows:

$$FAR = \frac{1}{N} \sum_{i=1}^{N} FAR(n)$$

(1.4)

$$FAR(n) = \frac{\text{Successful imposter attempts against person } n}{\text{Total imposter transactions against person } n}$$

(1.5)

Where $FAR(n)$ represents the FAR value for person $n$ and FAR is the average over all users in the database. Both FRR and FAR are closely related to the threshold value $T$. It is possible to achieve any value of one error rate at the expense of other. For example, in very high security applications it is desirable that no imposter attempts are accepted by the biometric system. This means $FAR = 0$. To achieve such a value for $FAR$, the threshold value needs to be increased. But when the threshold value is increased $FRR$ also increases with it thus rejecting many qualifying users. Figure 1.4 depicts the variation of FRR and FAR with the similarity measure $S$.

![Figure 1.4: Biometric System Errors](image)

The value of threshold, $T$, is tweaked to get the desired response from the biometric system. The value of $T$ can be adjusted such that both FRR and FAR are equal. This
value of error is called Equal Error Rate (EER).

\[ EER = FRR = FAR \] (1.6)

A plot between FAR and FRR at different values of \( T \) is called Receiver Operating Characteristic (ROC). An example ROC is shown in Figure 1.5.

![ROC curve showing relationship between FRR and FAR](image)

Figure 1.5: ROC curve showing relationship between FRR and FAR

The performance metrics discussed so far are mostly used in the verification systems. In the identification scenario, Cumulative Match Score (CMS) is commonly used to assess the system accuracy. CMS is defined by the following equation.

\[ CMS = \frac{Total \ correct \ matches}{Total \ probes} \] (1.7)

Biometric systems can also be used to short list the number of candidates that match closely to the probe ID. A human expert then makes the final decision from the short listed list. This can save considerable amount of time and resources as the human expert does not need to go through all the records. In this kind of set up the probe is compared
to the gallery gait sequences and then gallery is sorted in descending order according to the value of the similarity measure $S$. Then CMS value at different ranks is plotted to get Cumulative Match Curve (CMC). A CMS value at rank $r$ is defined below.

$$CMS_r = \frac{\text{Total correct matches in top } r \text{ of sorted list}}{\text{Total probes}}$$  \hspace{1cm} (1.8)

A typical CMC is shown in Figure 1.6.

![Figure 1.6: A Typical CMC](image.png)

1.5 Body Dynamics and Gait Recognition

The most critical step in gait recognition system is the extraction of gait features from video data. Human gait is cyclic in nature and this characteristic exhibits itself in cyclic appearance changes in the images when taken from a side view. Although gait is a dynamic process, studies have shown that static body parameters such as length and widths of limbs are also important in gait recognition. In appearance based methods, dynamics of lower half of the body are generally considered more important. Studies have been performed on the relative importance of static and dynamic features in gait
recognition. But there is a lack of work in relative analysis of dynamic features from different parts of the body especially in an appearance based set up. Dynamics of different parts of the human body play a role in characterizing the human gait pattern. This work analyzes the discriminatory power of features extracted from different parts of the body by applying area masks.

1.6 Contributions

This dissertation evaluates the performance of dynamic features extracted from different parts of the silhouettes. The relative significance of these dynamic features is established in an appearance based set up. This comparative study of dynamic features can help in better understanding of the walking process and gait determinants in biometric recognition. The results can be applied to design a more robust and reliable feature set for gait recognition. The movement of the lower half of the body is considered a major determinant of human gait. This work sheds some light on the importance of upper and lower arm movements in gait recognition. There are situations where it is not possible to extract quality features from leg movement due problems such as noise, shadows or occlusion. We show that lower arm and even shoulder movement may be used in those situations to extract gait signatures thus avoiding the noisy data from lower part of silhouettes.

We also propose two new feature extraction techniques for gait recognition. The first type of gait features are based on 2nd order and 3rd order correlation functions. Both new gait features are evaluated by comprehensive experimental analysis and their performance is compared with the current state of the art techniques. The proposed correlation based features are also combined with other existing features to enhance the performance of the system.

We also present another new feature extraction technique based on discrete wavelet
transform. The proposed gait feature is very compact and the computational complexity of the feature extraction is low compared to the other appearance based methods. This can significantly save computational resources during off line and online processing of the gait data. We show that the performance of the wavelet based features is not very robust and degrades significantly due to variations in the recording conditions.

We also show how the performance of the same technique varies when the data set is changed. The performance of the benchmark Baseline algorithm drastically degrades when it is implemented on a different database. We show that the nature of the database and silhouette sequences plays a role in determining the recognition performance of the gait features. We also show that the performance of the system can be enhanced by extracting dynamic gait features from certain parts of the silhouettes based on the comparative study of feature relevance.

The following papers published in the refereed journal and conferences are based on the contents of Chapters 3, 4 and 5.


The following journal paper has been submitted and is under review while conference paper is under preparation and will be submitted to an IEEE conference.

• T. Amin and D. Hatzinakos, “Gait Recognition by Multi-scale Features”, To be submitted to a conference.

1.7 Thesis Organization

The rest of the dissertation is organized as follows.

Chapter 2 presents a survey of related works in gait recognition. Research literature in both model based and appearance based categories is reviewed. The overview of works comparing static and dynamic features is also given in this chapter.

Chapter 3 introduces the experimental setup for performance evaluation. The gait databases used for experimental evaluation are described briefly in this chapter. The relative importance of dynamic features extracted from different parts of the silhouettes is established here. This chapter also discusses the similarity criteria and the feature vector normalization procedure adopted in this dissertation.

Chapter 4 presents the correlation based feature extraction from the silhouette sequences. Both 2nd. order and 3rd. order autocorrelation functions are used to extract gait features for recognition. The recognition efficiency of both features is compared with other state of the art techniques. This chapter also shows how the performance of the system can be enhanced by combining different type of features.

In Chapter 5, cyclic nature of human gait is captured by area signals and is then analyzed using wavelet transform. A comprehensive experimental evaluation using different wavelet kernels at different levels of decomposition is given in this chapter. The recognition performance of the proposed wavelet based gait feature is also compared with some existing techniques.
Chapter 6 presents conclusions of this dissertation and lists some potential extensions of present work in the future endeavours.
Chapter 2

Literature Review

Human recognition by gait was motivated by the psychological studies of the motion perception by Johansson [4]. Johansson used Moving Light Display (MLD) to study the motion perception. Light bulbs were attached to the person’s joints in MLD experiments. Subjects were then filmed performing different types of motion activity such as dancing, walking and running in dark background. These films only show the collection of bright spots in a two dimensional plane and carry no structural information because the bright spots were not connected. The images of the bright spots were shown to different observers and it was noted that these scattered spots were meaningless to them. But when the films were played, the movement of the spots created impressions of a person walking, dancing and running etc. This shows that the relative movements of certain joints in the human body carry information about personal walking styles and dynamics. The position of light bulbs and corresponding point display is shown in Figure 2.1 for walking and running movement. It was also noted that the familiarity of an observer with a particular type of motion plays an important role in its recognition. The inverted MLDs were not recognized by observers as a walking or dancing pattern.

C. D. Barclay et. al. showed that the identity of a friend and the gender of a person can be determined from the movement of light spots only [5]. A database of 7 male and 7
female walkers was used in their experiments. They investigated the temporal and spatial factors in gender recognition by point light displays. They showed that the duration of dynamic stimulus plays critical role in the recognition. They determined a threshold of two step cycles for this duration. In the spatial domain, the shoulder movement for males and hip movement for females were found to be important determining factors. A very interesting observation was made when upside down film was played to the observers. It had the effect of reversing the gender appearance making females look like males and vice versa.

The pioneering works on human motion analysis and gait recognition described in the preceding paragraphs fall into the category of marker based techniques. Human gait analysis and recognition techniques can be divided into the following two main classes.

- **Sensor Based**: Wireless or wired sensors are attached to the joints of the subjects and the displacement and angles are recorded with respect to a reference point. Joint angle trajectories and the distance between different parts of the body during gait is then calculated and used for recognition and analysis. Pressure plates or carpet is also used to measure the pressure profile of the feet during gait. Sensor based methods are more prevalent in medical research and rehabilitation studies. This is done in the laboratory setup and therefore has limited scope.

- **Image Based**: No sensors are attached to any part of the body but a video is recorded as the person walks along a preset trajectory. The data can be recorded
both indoor and outdoor using single or multiple cameras. This type has much wider scope and applications as even the existing video data and video data recorded for other purposes can be analyzed. Image based class can be further subdivided into two categories i.e. *marker based* and *marker-less*. Active or passive markers are placed on the body of the subject at different joints in marker based methods. This helps to detect and track the motion of desired joints in the video during gait motion. The subjects usually wear black and tight clothing and then reflective markers are placed on the joints. Markers of color different from the color of subjects’ clothing are also used to help in the tracking process. In the marker-less methods, a video is recorded without using any markers and the data is recorded with normal clothing.

This work falls into image based marker-less category. There are two main approaches to human gait analysis and recognition in marker-less systems. In the first approach, known as appearance based method, no *a priori* human geometric shape model is assumed. While in case of model based approaches, a priori geometric shape model is available. We will present a brief review of the past works in both categories with more emphasis on the recent developments.

### 2.1 Model Based Gait Recognition

Model based approaches assume *a priori* human shape/geometric model to analyze the motion and shape of human body parts. There has been considerable work on tracking human body based on the shape models during the past few years. However, model based techniques with particular focus to gait analysis have not caught much attention of the researchers’ community. This is partly due to the reason that tracking of human body is in itself a challenging problem involving very intensive computations. The geometrical model is usually parameterized and tracking of the shape is achieved by establishing the
correspondence between model configurations and image features. The most common methods for tracking include Kalman filter [6], dynamic Bayesian network [7] and condensation algorithm [8]. The model based approaches extract gait features from either static parameters or relative motion of joint angles. The static parameters such as torso height, leg length and stride are calculated from fitting the model in each frame and then further analyzing it for feature extraction. The joint angle trajectories are calculated in some methods and gait features are extracted from them. The model based approaches can also be distinguished by the dimension of the shape model. The shape model can be 2D or planar or a 3D model. The following paragraphs include a brief description of model based works in human gait analysis and recognition.

Niyogi and Adelson formed an XYT spatio-temporal cube by stacking each of the frames in an image sequence one right after another [9]. A unique braided signature of walking patterns extracted by the XT-slice of the cube near the walker’s ankle shows two legs criss-crossing over one another as the walker walks from left to the right. Their approach consists of finding translating blobs in image sequences, and testing if the XT-slice of the lower half of the blob contains a gait signature. After detecting the human walker by this gait signature, the spatio-temporal edges for all XT-slices in the translating blob are recovered. A stick model of the person is generated from these contours which are then used for the gait recognition based on certain assumptions.

Gavrila and Davis recovered 3D body pose at different time instants from a sequence of images acquired from multiple views [10]. They used *a priori* knowledge about the kinematic and shape properties of the human body to make the tracking tractable. The purposed model has 22 Degrees Of Freedom (DOF). They formulated the problem as a search problem of finding the pose parameters of a graphical human model whose synthesized appearance is the most similar to the actual appearance of the real human. Search space decomposition was used to overcome the problem of huge dimensionality (22D human). A novel similarity measure between the synthesized appearance and actual
appearance is also defined which is based on the whole contours/regions rather than a few points. The ambiguity and occlusion problem in a one view is resolved by using multiple views. A database using 4 views with subjects performing different motion activities such as hand waving and Tango was used for experimental evaluation. The work was performed in context to deriving better gait features but they did not report any recognition results. Only the tracking performance of the technique was reported in the paper. Wren et. al. developed a real-time system called Pfinder to track and interpret human behavior [11]. They used 2D model for tracking and detection of human body by Maximum a Posteriori (MAP) probability estimation. They modeled the human body as a collection of 2D blobs. These body blobs are described by spatial and color Gaussian distribution. The foreground is segmented from the background by using a background model and blobs representing human hands, head and feet etc. are then fit over the foreground region. The body parts were identified by using a 2D contour shape analysis. The system was used for gesture control, recognition of American sign language, creation of avatars and to establish tele-presence.

Deutscher, Blake and Reid modified the particle filter for the tracking of articulated body motion with a large number of DOF [12]. They called their implementation of the particle filter as annealed particle filter. The complexity of the search problem increases exponentially with increasing number of DOF. In order to decrease the number of samples required to propagate over time for tracking, they used the concept of simulated annealing to modify the particle filter. The posterior conditional probability distribution of input state variable is represented by samples along with their weights. A simpler weighting function was used instead of directly evaluating the posterior probability for each configuration of the state variable. The motion of a model with 29 DOF was tracked with considerably less number of particles than the original condensation algorithm. The experimental evaluation showed a better performance than the condensation algorithm by using this strategy. In [13], Huang et. al. presented a method for human body tracking
based on the 2D model. Their 2D card board model is the extension of the 2D scaled prismatic model with one additional DOF for the width change. They used a mixture model to represent the movement of the body. The motion parameters of the articulated body motion are solved using the Expectation Maximization (EM) algorithm.

Four static body parameters were used for gait recognition on a database consisting of 15-18 subjects by Bobick et. al. [14]. The extracted parameters were: vertical distance between the head and foot, distance between the head and pelvis, the distance between the foot and pelvis and the distance between the left foot and right foot. The distances were measured in number of pixels and a depth compensation mechanism was used to convert from image to world units. The gait feature vector is very compact but the recognition performance is low for rank 1. Yam et. al. developed an automated technique capable of recognizing people from the walking as well as from running gait [15]. They used a modeling technique based on the concept of coupled oscillators and the underlying principles of human locomotion. The two approaches given in their paper derive a phase-weighted Fourier Descriptor (FD) gait signature by automated non-invasive means. Assuming the gait symmetry, the same model was used to describe either leg since both perform the same motion but out of phase with each other by half a period. These motions operate in space and time satisfying the rules of spatial symmetry and temporal symmetry. Both legs were modeled by two distinct coupled oscillators oscillating at the same frequency but with a phase difference. This model of forced coupled oscillators is fitted to the image data extracting the lower leg motion in both walking and running gait. The gait features were derived from the magnitude and phase of FDs of thigh and lower leg rotation. A statistical analysis was also performed to find the most effective feature set.

Green and Guan defined the alphabet of human movement called dynemes which are the smallest units of motion [16, 17]. The combination of these units in different order forms different skills and activity. They developed a 3D clone body model which
is dynamically sized and texture mapped to each person enabling both edge and region tracking. A particle filter with forward smoothing is used for the estimation of the parameters of the body model which has 32 DOF. The gait signature was extracted by using the Fourier series to describe the leg motion. The method was tested on a database of 58 people walking in a sagittal plane wearing tight fitting clothes. The training set consisted of 48 people while the additional 10 people were used for the testing. They achieved an accuracy of 88% by using the extracted gait signatures. The anthropometric features were also tested to evaluate their performance in human recognition. Interestingly, a recognition rate of 92% was achieved which is 4% higher than the gait based analysis. A combination of both gait and anthropometric features increased the accuracy to 94%.

The method proposed by Raquel Urtasun and Pascal Fua is based on the fitting of 3D temporal motion models to synchronized video [18]. They not only achieved tracking by this method but also recovered motion parameters which were then used for human recognition. They formulated tracking problem as a minimization of differentiable objective functions whose state variables are the Principal Component Analysis (PCA) weights. The differential structure of these objective functions takes the advantage of standard deterministic optimization methods whose computational requirements are much smaller than those of probabilistic ones.

In [19], a 3D articulated body model defined by 16 links and 22 DOF was used with certain constraints on the movements of arms and legs to reduce the complexity. The motion trajectory of the walker’s footprints is detected from the segmented video sequence. 3D human model is then moved on this trajectory driven by the prior motion model and the joint angles are adjusted to the walking style. The extracted joint angles were not used for recognition and only gait analysis was performed. A statistical model for detection and tracking of human silhouette and the corresponding 3D skeletal structure in gait sequences was proposed by Carlos et. al. [20]. A different point distribution model was applied depending on pose. The performance of the model is improved by
taking into account temporal dynamics to track the human body. The incorporation of temporal constraints on the model helps increase the reliability and robustness. The 3D skeletal structure is extracted and tracked over time in the image sequence. Wagg and Nixon developed a new model-based method based on the biomechanical analysis of walking people and used it for recognition [21]. The image sequences were segmented to extract the moving regions and an articulated model is fitted to the edge by a hierarchical procedure. Motion estimation is performed by using a sinusoidal model for the leg and angle trajectories are extracted. The method is evaluated by using SOTON database and the feature vector is 63 dimensional. A recognition rate of 84% on the indoor dataset and 64% for the outdoor dataset was achieved.

Lu et. al. proposed a layered deformable 2D body model for gait recognition [22]. Their model is a full body model consisting of 10 body segments specified by 22 parameters. These 22 parameters define the size, position and orientation of the body segments. The limb orientation and position was estimated using mean shift algorithm for manually labelled silhouettes. The joint angles were then calculated from limb orientations and positions using simple geometry. A coarse to fine estimation based on the ideal human body proportions (eight-head height) was proposed for automatically extracted silhouettes. DTW was used for pattern matching between gallery and probe sequences. The performance of the features for gait recognition was not very impressive and it further degraded when automatic silhouettes were used. The average rank 1 performance of 25% was achieved for manually extracted silhouettes which dropped to 18% for the automatically calculated silhouettes. A 3D human body model consisting of 11 body segments was developed by Gu et. al. [23]. The head was represented by a sphere and other segments were cylindrical. The model contains 10 joints with 24 DOF. The kinematic structure of the model was estimated by employing anthropometric constraints between ratios of limb lengths. After the body segmentation, adaptive particle filter was used to track the body segments. Gait features were extracted from pose parameters and joint position
sequences. Two gait models were obtained from normalized joint sequence of the whole body and the normalized joint sequence of two legs using an exemplar-based Hidden Markov Model (HMM). MAP estimation was used for pattern classification. The test database consisted of multiple video streams of 12 subjects that were simultaneously captured from multiple static calibrated cameras. Volumetric representation sequences were created using visual hull method after foreground extraction. An average recognition rate of 94.4% was reported on the test database.

In [24], Arai et. al. reported gait recognition results on Chinese Academy of Sciences (CASIA) data set consisting of 31 male and 31 female subjects. They extracted silhouettes by simple background subtraction. The skeleton was then extracted using thinning and other morphological operations. Eight important feature points were then determined on the extracted skeleton structure. The skeleton was reconstructed by connecting 8 points with straight lines. Motion was also estimated using simple frame subtraction method. Discrete wavelet transform was used on skeleton data and motion signals to extract features for recognition. They achieved an average correct recognition rate of 95.97% on the test database. Table 2.1 summarizes and compares the performance of model based gait recognition research works.

### 2.2 Appearance Based Gait recognition

The appearance based gait recognition methods first perform motion detection to segment the regions corresponding to the moving humans. Some form of shape analysis is then applied to these human image sequences to extract the gait signatures. Static body parameters such as lengths and widths of limbs, height of the person are extracted in some techniques and used to represent gait. Some works rely on the dynamic features that are extracted by shape changes and motion flow. Majority of the techniques in
Table 2.1: Model based gait recognition research

<table>
<thead>
<tr>
<th>Authors</th>
<th>Body Model</th>
<th>Database Size # Subj./Seq.</th>
<th>Database Complexity</th>
<th>Average Performance(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niyogi et. al. [9]</td>
<td>Stick Figure</td>
<td>5/26</td>
<td>Low</td>
<td>81</td>
</tr>
<tr>
<td>Gavrila et. al. [10]</td>
<td>3D</td>
<td>3/3</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>Deutscher et. al. [12]</td>
<td>2D</td>
<td>1/1</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>Huang et. al. [13]</td>
<td>2D</td>
<td>1/1</td>
<td>low</td>
<td>NA</td>
</tr>
<tr>
<td>Bobick et. al. [14]</td>
<td>2D</td>
<td>15/268</td>
<td>Medium</td>
<td>87.78</td>
</tr>
<tr>
<td>Green et. al. [17]</td>
<td>3D</td>
<td>58/58</td>
<td>Low</td>
<td>94</td>
</tr>
<tr>
<td>Urtasun et. al. [18]</td>
<td>3D</td>
<td>1/1</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>Sappa et. al. [19]</td>
<td>3D</td>
<td>1/1</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>Wagg et. al. [21]</td>
<td>2D</td>
<td>115/4824</td>
<td>Medium</td>
<td>72.95</td>
</tr>
<tr>
<td>Haiping et. al. [22]</td>
<td>2D</td>
<td>72/287</td>
<td>High</td>
<td>25/18 (^1)</td>
</tr>
<tr>
<td>Gu et. al. [23]</td>
<td>3D</td>
<td>12/12</td>
<td>Low</td>
<td>94.1</td>
</tr>
<tr>
<td>Arai et. al. [24]</td>
<td>Stick Figure</td>
<td>62/264</td>
<td>Low</td>
<td>95.97</td>
</tr>
</tbody>
</table>

appearance based category work on human silhouette sequences.

In [25], Liu and Picard, presented an algorithm for simultaneous detection, segmentation, and characterization of spatiotemporal periodicity. The algorithm may be applied to find the periodicity in the images which can be used for the object detection and classification. The work is motivated by the notion that the human gait/motion is periodic. This characteristic has been used to separate and segment different objects from the images such as dogs and cars were separated from the human objects. The proposed algorithm acts as a periodicity filter and is computationally simple. It was shown to be more robust than optical flow based techniques in the presence of noise.

Meyer et. al. modeled several body parts such as head, trunk and leg as well as the

^1manual/automatic silhouettes
background as mixture densities in grey scale images [26]. They localized the body parts in every frame by mixture densities and accounting the anatomic relationships between the body parts. The mixture model is calculated using EM algorithm. Features are extracted from the trajectories and HMMs are trained. One HMM represents each kind of gait such as walking, running and dancing. Cutler and Davis developed a system for object classification based on the self-similarity of the object during motion [27]. The self similarity of the human objects in the images shows a periodic variation because of the periodic nature of human gait. The algorithm developed by them consists of two parts. In the first part, the object of interest is segmented from the background based on motion. In the second part, the self similarity of the object is computed as it moves in time. A time-frequency approach is then applied to analyze the periodicity of the self-similarity plots.

Stauffer and Grimson adopted a novel approach for the motion tracking [28]. They modeled the values of each of the pixels with mixture of Gaussians rather than modeling the values of all the pixels with a particular type of distribution. Based on the persistence and variance of each of the Gaussians of the mixture, they determined the Gaussians corresponding to the background colors. All the pixel values that do not fit the background distributions form the foreground. They used an on-line K-means approximation to update the model because EM algorithm would be very costly in such a situation where each of the pixel value is modeled with a Gaussian distribution. The foreground pixels are then segmented into regions by applying a two-pass connected components algorithm. After motion tracking, they classified the silhouettes and detected the unusual events. Shutler, Nixon, and Harris used statistical analysis by using temporal moments [29]. They proposed velocity moments based on the center of mass. Background was extracted using the temporal mode filter. The subjects were extracted by selective subtraction and region growing. The velocities were then calculated using the dense optical flow fields. The average velocity of each person was used to calculate the
velocity moments up to the fourth order. They clustered the velocity moments to show that each subject forms a distinctive cluster.

Symmetry is a fundamental principle and most of the objects exhibit some form of symmetry. James et al. proposed to use the symmetry of motion to distinguish between human and animal motion [30]. The symmetry information was estimated from the images using generalized symmetry operator which assigns a symmetry measure to each point in the image. They reported a recognition rate of 100% using silhouettes from SOTON (University of Southampton) database consisting of 16 sequences from 4 subjects. The recognition rate for University of California, San Diego (UCSD) database was slightly lower at 97.6% obtained by using silhouette data. The UCSD database used in their experiments consists of 42 sequences from 6 subjects.

BenAbdelkader, Cutler and Davis contended that planar dynamics of a walking person are encoded in 2D similarity plots between pair of images taken from the sequence of the walking person [31]. Assuming that camera is sufficiently far from the moving person, the camera projection becomes approximately orthographic with scaling. Under the orthographic projection and if the motion of the points are constrained to the planar motion, then the object dynamics are completely preserved in the projection up to a scale factor. Taking these assumptions, they first segmented the moving person from the background. The image templates were then scaled to the uniform height because the sizes may vary due to the depth variations and segmentation errors. A self-similarity plot is then obtained by correlation of each pair of images in the sequence. They used PCA to reduce the dimensionality of the input feature space. A recognition rate of 12% was achieved on Carnegie Mellon University (CMU) MoBo dataset when training on slow-speed sequences and testing with moderate-speed sequences. However, the recognition rate was increased to 72% when both the training and testing was performed using slow sequences. A recognition rate of 76% was reported for the fast sequences.

To benchmark the performance of gait recognition techniques, a Baseline algorithm
was presented in [32] by Philips et. al. The Baseline algorithm used the correlation
between the silhouettes as a feature to represent gait. The gait sequences were segmented
and the similarity measure based on the maximum correlation between the gallery and
probe silhouettes was used. The paper also described the Gait Challenge (GC) database
in detail and seven probe sets were given to assess the performance in different recording
conditions. In [33], BenAbdelkader et. al. presented a parametric method for personal
identification based on the height and stride parameters of the gait from low resolution
video sequences. A non-parametric background modeling approach was adopted for the
segmentation of the moving objects. Foreground blobs were then tracked using spatial
and temporal coherence. The height and stride parameters were determined from the
extracted binary silhouettes. The experiments were performed on a database containing
45 subjects and an accuracy of 49% was achieved by using both the stride and height
parameters and only 21% by using the stride parameter only. Although, they didn’t
achieve a significant performance, yet their results show that stride and height parameters
may be used as potential candidates for the gait recognition systems.

Sunderesan, Chowdhury and Chellappa developed a general framework for recognition
of humans using gait [34]. This framework is based on the HMM model. The framework
assumed that the individual transitions among N discrete states during a walk cycle. An
adaptive filter was used to detect the cycle boundaries. The framework is independent
of the feature vector and can be adapted to different feature sets. The statistical nature
of the HMM makes the model robust. They used binary images of the foreground after
the background subtraction as feature vector. The experimental evaluation was done
using GC database consisting of 75 subjects and 7 different probe sets. They achieved
99% CMS for probe A which dropped to 18% for probe G. The variation in results using
different similarity measures was also reported.

Wang et. al. developed a silhouettes based recognition system in [35]. They mod-
eled the background using least median squares method from a small portion of image
sequences including the moving object. The segmentation was performed using the differencing between the computed background and the current image using a heuristic function. The process was performed on each color channel (Red, Green and Blue) in the image. The pixels determined as changing point by any one of the color channel were labelled as foreground pixels. Morphological techniques of erosion and dilation were used to filter the spurious pixels and to fill small holes inside the silhouettes. The contours of the silhouettes were extracted using a contour following algorithm. The centroid of the contour was calculated and 2D contour was converted to 1D distance function by calculating the distance of each point from the centroid. PCA was trained to represent the features in a low dimensional eigenspace. The classification was performed by the nearest neighborhood classifier using normalized Euclidean distance and spatiotemporal correlation.

In [36], Foster et. al. extracted the silhouettes by applying the chroma-key subtraction in conjunction with a connected component algorithm. After getting the silhouettes, they applied different area masks to the images and calculated the area under these masks. The area history for the different masks was thus obtained which carry the gait information. The area vectors from all the masks were then concatenated together and form the gait feature vector. Experiments were performed on the SOTON database. It was observed that the area vectors relating to the horizontal masks gave much higher discrimination than the vertical masks. They achieved a recognition rate of 76.6% by combining all the area vectors.

In [37], Liu and Sarkar computed the average silhouette during the whole gait sequence. Their algorithm consists of three steps. In the first step, the background pixel statistics were calculated using the Mahalanobis distance and EM algorithm. The second step calculated the periodicity of the gait by simply counting the number of foreground pixels in the silhouette in each frame over time. The pixels belonging to the leg area were used to increase the sensitivity for determining the periodicity of the gait. In the
third and last step, the average silhouettes were computed. The similarity measure is defined as the negative of the median of the Euclidean distance between the averaged silhouettes from the probe and the gallery. Han and Bhanu calculated gait frequency and phase from the lower part of the silhouettes using silhouette sequence data [38]. They estimated phase and frequency by maximum entropy spectrum estimation to avoid the problem of side lobes due to noise. They also calculated a grey level gait energy image from the silhouette sequences. PCA followed by multiple discriminant analysis was applied to extract features from the real and synthetic gait energy image templates. Different features were fused together and experimental results showed a marked improvement of recognition results over the Baseline algorithm. Yu et. al. used FD for human gait recognition [39]. They eliminated the influence of the walking speed by using the same number of silhouettes in each gait cycle. The experimental results showed a recognition rate of 85.2% at rank 1 for the key FD features.

Boulgouris, Plataniotis and Hatzinakos developed a new methodology for gait recognition based on the concept of DTW [40]. They exploited the periodicity of walking to partition the gait sequence into cycles. They located the frame indices where the sum of the pixels is minimized corresponding to the half gait cycle. Autocorrelation function of the sum of the pixels was used to determine the cycle length because of the noisy nature of the original function. DTW was performed between all cycles of probe and gallery sequences after the partitioning of sequences into gait cycles. The recognition rate using this method showed increased performance over the Baseline algorithm on the GC database. They also developed an angular transform which gives the average distance from the centroid of the silhouette to a small group of pixels on the contour of the silhouette [41]. The silhouettes were pre-processed before the application of the angular transform to remove isolated errors and artifacts. All the silhouettes were aligned so that their centers were at the centre of respective frames. The method was evaluated using the gait challenge database and performed considerably better than the Baseline
algorithm. Linear Time Normalization (LTN) with angular transform features was used in their later work reported in [42]. They found that linear time normalization worked better than DTW and succeeded in achieving 8-20% increase in recognition performance with compared techniques on GC database. The authors also studied the recognition performance of 4 sets of gait features using 5 different recognition methods [43]. An average CMS of 36% was obtained by using DTW, LTN, HMM and structural matching at rank 1. The performance at rank 5 was about 60% for all these 4 pattern recognition techniques. The only exception was noted using the frequency domain distance which performed significantly lower achieving only 20% and 41% average CMS value at rank 1 and rank 5 respectively.

A subspace approach based on the matrix representation of gait data was proposed by Xu at. al. [44]. Traditionally, the image matrix is concatenated into a single dimensional vector to apply PCA and Linear Discriminant Analysis (LDA). The well known curse of dimensionality due to large dimension compared to much smaller number of samples give rise to errors. The proposed matrix based coupled subspace analysis and discriminant analysis with tensor representation is an attempt at resolving the dimensionality issue. They achieved a significant performance increase over the Baseline algorithm. A CMS of 89% was achieved for probe A at rank 1 compared to 73% for the Baseline algorithm. Ioannidis et. al. designed three new feature extraction methods for gait recognition [45]. Two methods described as radial integration transform and circular integration transform are based on radon transform. Their third approach for feature extraction was based on weighted Krawtchouk moments. The depth information can also be incorporated if available. The recognition results were the highest for the Krawtchouk moments followed by radial integration transform and circular integration transform on GC database. They also used a feature fusion scheme based on genetic algorithm to improve the recognition performance. An improvement of 1-8% was obtained using all three types of features.

Boulgouris and Chi used Radon transform of the binary silhouettes to generate tem-
plates. LDA and subspace projection was applied to obtain a low-dimensional feature vector consisting of selected Radon template coefficients. These selected Radon features were used for gait based recognition [46]. Each gait sequence was represented by a sufficiently compact signature of 40 coefficients. A significant improvement in recognition performance was achieved by them over the Baseline benchmark results. Tao et. al. developed general tensor discriminant analysis to apply directly to tensor data without any vectorization [47]. This results in overcoming the under-sampling problem and also provides more robust features. Each gait sequence was partitioned into gait cycles after determining the gait period. The silhouettes in each gait cycle were averaged to obtained one image that represented the whole cycle. these images were then used as features directly for classification using general tensor discriminant analysis. Gabor based features were also used in combination with general tensor discriminant analysis. The highest average CMS of 60.58% was achieved for a combination of Gabor, general tensor discriminant analysis and LDA.

Yang et. al. decomposed gait energy image using Gabor wavelet kernels with 5 different scales and 8 orientations [48]. The feature vector was constructed using the Gabor phase and LDA was applied to reduce the dimension of the feature space. Comparative performance on GC database showed that Gabor phase possesses more discriminatory power than Gabor magnitude. An average CMS of 62.25% was achieved using Gabor phase compared to 51.88% for magnitude. Multilinear PCA (MPCA) was developed by Lu et. al. in [49]. MPCA was introduced to apply on the 3D gait data directly by representing it as tensors. The application of subspace projection directly to 3D gait tensor data mitigates the famous curse of dimensionality problem. It also preserves structural information which is lost when data is vectorized for processing with traditional PCA and LDA. The tensor data was first normalized to make all tensors equal dimension. MPCA is then applied to obtain eigentensors. Classification was performed using different distance functions. The GC database was used for performance evaluation. The average
recognition performance of 54% and 76% was obtained at rank 1 and rank 5 respectively. This is a significant improvement over Baseline algorithm performance of 42% and 79% at rank 1 and rank 5 respectively. In a latter work [50], they introduced uncorrelated multilinear discriminant analysis which was shown to perform even better than MPCA.

Chen et. al. proposed a layered time series model which is a two level model combining HMM and dynamic texture model [51]. The gait cycle was first partitioned into temporally adjacent clusters of equal number of frames. Frieze feature and wavelet feature were then extracted from these clusters. Individual linear dynamic texture models were trained for each cluster that represent the states of the HMM. The evaluation was done using CASIA gait dataset B consisting of 124 subjects recorded from 11 views [52]. Wavelet features outperformed the frieze features in their experimental analysis. An average recognition rate of 95.7% was obtained using layered time series model technique which was higher than that of dynamic texture model and HMM results of 58.6% and 93.9% respectively. Wang et. al. modified the gait energy image and constructed chrono gait image to include temporal information [53]. After gait period detection, they used local information entropy to obtain the gait contour from the silhouette images. Synthetic chrono gait images were also constructed to avoid over fitting due to smaller number of real chrono gait images. LDA and PCA were applied for dimensionality reduction. A comprehensive experimental evaluation was reported using 3 major gait databases. An average CMS value of 48.64% and 66.81 was achieved at rank 1 and rank 5 respectively using all 12 probe sets of GC database. These results did not show marked improvements over related gait energy image method and were only marginally higher. Table 2.2 summarizes and compares the performance of appearance based gait recognition techniques.
Table 2.2: Appearance based gait recognition techniques

<table>
<thead>
<tr>
<th>Authors</th>
<th>Database Size</th>
<th>Database Complexity</th>
<th>Average Performance(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>James et. al. [30]</td>
<td>6/42</td>
<td>Low</td>
<td>97.6</td>
</tr>
<tr>
<td>Philips et. al. [32]</td>
<td>71/452</td>
<td>High</td>
<td>35.92</td>
</tr>
<tr>
<td>BenAbdelkader et. al. [33]</td>
<td>44/176</td>
<td>Low</td>
<td>77</td>
</tr>
<tr>
<td>Chellappa et. al. [34]</td>
<td>71/452</td>
<td>High</td>
<td>53.51</td>
</tr>
<tr>
<td>Foster et. al. [36]</td>
<td>28/112</td>
<td>Medium</td>
<td>75</td>
</tr>
<tr>
<td>Liu et. al. [37]</td>
<td>71/287</td>
<td>High</td>
<td>23.48</td>
</tr>
<tr>
<td>Boulgouris et. al. [43]</td>
<td>71/452</td>
<td>High</td>
<td>36</td>
</tr>
<tr>
<td>Xu et. al. [44]</td>
<td>71/440</td>
<td>High</td>
<td>55.48</td>
</tr>
<tr>
<td>Lu et. al. [49]</td>
<td>71/452</td>
<td>High</td>
<td>54</td>
</tr>
</tbody>
</table>

2.3 Miscellaneous Methods

The traditional approach to human gait recognition is to construct some form of feature template and then apply pattern recognition methodology to compare these templates. The features are generally extracted from the silhouette sequences recorded from the side view using single camera. There are some works reported in literature that have departed from this traditional paradigm. We will discuss a few of those representative works in the preceding paragraphs.

A channel coding method based on distributed source coding principles was adopted for human gait recognition by Argyropoulos et al. [54]. The framework is different from the traditional pattern recognition approach that is used for feature matching in gait recognition works. They experimented with features extracted using radial integration transform, circular integration transform and Krawtchouk moments. The gait features are then coded using SlepianWolf encoder implemented by using a systematic low density
parity check encoder. In the authentication stage, the decoder decodes the codewords using belief propagation. The correct codeword is output if the error in the probe codeword is within the error correction capability of the decoder and the identity is verified. A performance gain of 10% to 30% was achieved for all the experiments compared to the Baseline algorithm. In [55], Shakhnarovich, Lee, and Darrell developed a view normalization method for multi-view integrated face and gait recognition. The technique involves the computation of an image-based visual hull from a set of monocular views. This computed visual hull is then used to construct virtual views for tracking and recognition. Canonical viewpoints are constructed by examining the 3-D structure, appearance, and motion of the moving person. The centroid of the silhouette is determined and used to divide the whole silhouette into 7 regions. An ellipse is then fitted to each of the regions and features are extracted from these regions in every frame. The mean and standard deviation of these features over time are collected together to form the gait feature vector.

The idea of using more than one view to extract gait features has been attempted in several works such as the one reported in [56]. The use of frontal view video instead of the usual side view for gait recognition was also reported in some works. In [57], Goffredo et al. used frontal view camera image sequence for gait recognition. After calculating the gait period, 3D gait volume was constructed using silhouettes from one gait cycle. The feature vector composed of 3D central moments of the gait volume and some scale dependent gait features including the number of frames for one gait cycle and the silhouette’s height and width maximum increment. They performed experiments on 3 data sets including CASIA-A and CASIA-B data. A correct classification rate of 91% and 97.92% was obtained for CASIA-A and CASIA-B respectively. Frontal view gait recognition was also tried in research reported in [58].

The quality of silhouettes is very critical for any gait recognition system and it is very hard to extract robust features from noisy silhouettes [59]. Liu and Sarkar performed silhouette reconstruction to remove noise and shadows. A population HMM was trained
using manually specified silhouette data. The states of HMM represent the stances and the transition probabilities capture the motion dynamics between the stances. Statistical shape model called the eigen-stance gait model was constructed for each stance using manual silhouettes. Each frame was later matched with these stance subspaces using the already learned population HMM. The silhouettes were reconstructed by projecting it on the matched eigen-stance model. The experimental evaluation showed that the performance of the Baseline algorithm actually dropped when manual silhouettes were used. This surprising result indicates that the quality of the silhouettes does not always explain the drop in performance especially in case of variation in surface and time between probe and gallery sets.

2.4 Dynamic versus Static Features

Human locomotion has been widely studied by medical and physiological research community. Their main purpose is to determine the gait variations and distinguish between normal and pathological gait and rehabilitation of the patients. Saunders et. al. have defined human walking as the translation of the center of mass of the body from one point to another in a way that requires the least energy [60]. They also identified the six gait determinants or variables that affect the energy expenditure. The six gait determinants are pelvic rotation, pelvic tilt, knee flexion at mid-stance, foot and ankle motion, knee motion and lateral pelvic displacement. The focus of this work as well as other similar ones was to study the movement of different muscles and limbs during the gait process. These types of studies are useful for detecting abnormalities in human walking and may also serve as a general guideline for recognition systems. The perception of human gait as well as its recognition involves much more than just the six determinants given in [60].

Das et. al. investigated the relative role of the temporal and spatial features using PCA [61]. The gait data was collected using the motion capture system with 13 markers
on the joints of human subjects. Their experiments showed that the temporal components determine the phase of the gait and account for approximately 70% variation in the data. However, the temporal components cannot distinguish the type of gait such as running and walking. The spatial components provide the cues to distinguish between running and walking. Wang et. al. used both static and dynamic body features for human recognition [62]. The static body features were derived from using the Procrustes shape analysis to obtain a compact appearance representation. The dynamic descriptors were estimated by recovering joint angle trajectories of the lower limbs using Condensation algorithm. The algorithm was evaluated by using a database consisting of 80 sequences from 20 subjects and four sequences per subject. They reported recognition of 83.75% at rank 1 by using only static features and a success rate of 87.5% when dynamic features were used. The combined features resulted in an increased recognition rate of 97.5% at rank 1. Wang et. al. also performed another comparative study between the dynamic and static features in their work reported in [63]. Their work showed that the dynamic information extracted from the video sequences is somewhat better for human identification than the static information.

Veeraraghavan et. al. conducted a detailed comparison between shape and kinematic features for human recognition [64]. Their experiments indicated that shape of the body carries more information than the kinematics for the recognition of humans from video sequences. However, using kinematics in conjunction with the shape features considerably improved the performance of the system. Similarly, gait analysis work carried out by Green and Guan also showed that anthropometric (static) features extracted by them were more discriminatory for human identification than the dynamic features in the shape of joint angle trajectories [17]. On the other hand, the experiments conducted by Johansson established the importance of dynamic features for identification [4].
2.5 Summary

In this chapter, we provided a brief overview of representative works in model-based and appearance-based gait recognition. A comprehensive survey of gait recognition techniques can be found in [65]. The problems and challenges in gait recognition signal processing were systematically described in [43] by Boulgouris et al. It is noted that in certain cases simpler techniques have produced much better results than those achieved by a lot complex and sophisticated methods. The quality of data and noise has always been considered a culprit responsible for errors and low efficiency. It was interesting to note that in one detailed study, it was found that the recognition results actually dropped when cleaner silhouettes were used. This may be the result of other variables affecting the performance of the system. Contradicting results were reported about the importance of dynamic and static features in gait recognition. This indicates the dependability of techniques on the database. The GC database has resolved this issue partially by providing a standard data set and a set of defined experiments for common framework of comparison. The dynamic feature comparison has not been performed explicitly in an appearance-based setup. We will shed some light on the dynamic feature performance extracted from different parts of the binary silhouettes in the next chapter.
Chapter 3

Determinants in Gait Recognition

In appearance based methods, the dynamics of the lower half of the body are generally considered more important. Studies have been performed on the relative importance of static and dynamic features in gait recognition. But there is a lack of work in relative analysis of dynamic features from different parts of the body especially in the appearance based set up. The dynamics of different parts of the human body play a role in characterizing the human gait pattern. In this chapter, we will analyze the discriminatory power of features extracted from different parts of the body by applying area masks. This will serve as a foundation for the selection of dynamic features and development of gait recognition algorithms in later part of this dissertation.

3.1 Extraction of Body Dynamics

Human walking process is cyclic in nature. The gait cycle is the time between two identical events during the human walking and is usually measured from heel strike to heel strike of one leg. A complete gait cycle is shown in Figure 3.1. The movement of arms and legs is the most prominent motion during the gait cycle. Assuming that image plane is perpendicular to the direction of motion, the gap between the two legs in 2D human silhouettes changes during the gait cycle. Similarly the gap between the arms and
the rest of the body also changes in a cyclic fashion. This dynamic information can be captured by applying area masks at different parts of the binary silhouettes similar to the approach adopted by Foster et. al. [36]. The number of pixels of the binary silhouettes under these masks is calculated. The process is repeated for each binary silhouette in the gait sequence and we obtain six area signals of length $N$, the number of frames in the gait sequence. The width of each area mask is 15 pixels. Figure 3.2 shows the location of six area masks for an example silhouette from the GC database. The following equations summarize the extraction of six area signals from the masks shown in Figure 3.2.

$$b[i, j] = \begin{cases} 
1 & \text{if pixel } [i, j] \text{ belongs to foreground} \\
0 & \text{if pixel } [i, j] \text{ belongs to background} 
\end{cases} \quad (3.1)$$

$$m_p[i, j] = \begin{cases} 
1 & \text{if } J_p \leq p < J_p + 15 \\
0 & \text{otherwise} 
\end{cases} \quad (3.2)$$

$$a_p[n] = \sum_{i,j} b_n[i, j] m_p[i, j] \quad (3.3)$$
where \( b[i, j] \) is the binary silhouette, \( m_p[i, j] \) is the area mask and \( a_p[n] \) is the area under mask \( p \) for frame \( n \) of the silhouette sequence. \( J_p \) is the starting row for mask \( m_p \) and \( p = \{1, \ldots, 6\} \) is the mask index.

These area signals are shown in Figures 3.3 and 3.4 for two typical silhouette sequences from the database. The area signals extracted by applying the area masks are noisy due to the imperfections in the silhouette extraction process. It is observed that a high frequency riding wave is present in all of the area signals. We apply a newly proposed Empirical Mode Decomposition (EMD) algorithm to remove these riding waves to get cleaner area signals [66]. The traditional data analysis methods such as Fourier transform have an inherent restriction to their application. They are suitable when the system is linear and the data is stationary. In most of the practical application scenarios these two conditions are rarely satisfied. But these traditional methods are still widely used because of their simplicity and well formed theoretical basis. In some cases the non linear and
Figure 3.4: Area signals for another silhouette sequence

Non-stationary data can be transformed to linear and stationarity data before processing it with Fourier based methods. But in many cases new methods are needed which can analyze non-linear and non-stationary data. There has been some progress in the analysis of non-stationary data in recent years. Wavelet analysis and Wagner-Ville distribution are the examples of data analysis tools for non-stationary data. EMD decomposes non-linear non-stationary data into oscillatory modes called Intrinsic Mode Functions (IMF) [66]. EMD algorithm is described in Appendix A.

In order to illustrate the noise removal by EMD, we choose two area signals from Figures 3.3 and decompose them using the EMD algorithm. The input signals are plotted in Figure 3.5 and their IMFs in Figures 3.6 and 3.7. The high frequency noise appears as the first IMF, $IMF_1$ as shown in Figures 3.6 and 3.7. The area signals are reconstructed by ignoring $IMF_1$ as given by the following equations.

$$\hat{a}_1[n] = \sum_{i=2}^{5} e_i[n] + r_k[n]$$

(3.4)
Chapter 3. Determinants in Gait Recognition

3.2 Correlation Analysis

After the noise removal from area signals by EMD algorithm, we compute autocorrelation of all six reconstructed area functions as follows:

\[
R_{\hat{a}_p}[l] = \sum_{n} \hat{a}_p[n] a'_p[n + l] \tag{3.6}
\]

Where \(R_{\hat{a}_p}\) represents the autocorrelation function of the reconstructed area signal \(\hat{a}_p\) and \(l\) is the time lag. \(R_{\hat{a}_p}\) is only calculated for positive lags i.e. \(l = \{0, 1, \cdots, N - 1\}\). The dynamic gait features are then derived by taking the Discrete Cosine Transform (DCT) of the six autocorrelation functions. The DCT of a discrete function \(R_{\hat{a}_p}\) is defined below.

Figure 3.5: Noisy area signals

\[
\hat{a}_3[n] = \sum_{i=2}^{5} e_i[n] + r_k[n] \tag{3.5}
\]

The reconstructed signals are shown in Figure 3.8.
Figure 3.6: Input signal $a_1$ and its IMFs

\[ T_p[k] = c[k] \sum_{l=0}^{N-1} R_{a_1}[l] \cos \left( \frac{\pi(2l + 1)k}{2N} \right) \]  
(3.7)

Where $T_p[k]$ is the DCT transform of the original signal $R_{a_1}[l]$ of length $N$. The coefficient $c[k]$ is given by:

\[ c[0] = \sqrt{\frac{1}{N}}, \quad c[k] = \sqrt{\frac{2}{N}} \]  
(3.8)

for $1 \leq k \leq N - 1$.

### 3.3 Experimental Setup

#### 3.3.1 Databases

There are several databases that were used for experimental evaluation of gait recognition works in the past two decades. The majority of databases were recorded indoors and the
number of subjects were small. Some examples of gait databases are MIT database, CMU MoBo database, SOTON database, Georgia Tech (GTech) database and GC database. Table 3.1 gives a summary of different databases used in gait based identification works. We use two gait databases in our experiments namely GTech and May 2001 GC database. The GTech database is a smaller database consisting of only 15 subjects. The May 2001 version of GC database is much bigger and contains 71 subjects. The GC database is the most comprehensive database among the ones mentioned above. Its full version consists of 122 subjects with 1870 sequences and six covariates. The details about these two databases are given below.

**GTech Database**: This database consists of sequences of 14-18 persons recorded using single camera with its viewing plane perpendicular to the ground plane. Multiple video sequences of the subjects were recorded indoor at two different camera angles using digital miniDV camera. The walking trajectories for two viewing angles were
The same subjects were then recorded outdoor after 6 months at two viewing angles. This time the viewing angles were fronto-parallel and 55°. The video data is available in .avi format and each file is about 25MB in size. The silhouette sequences of 15 subjects are also available for gait experiments [68]. The number of silhouette sequences in the database is 268. We use 5 probe sets for performance evaluation in our experiments. The first probe set consists of all the sequences in the database except the gallery sequences. The rest of the 4 probe sets has been defined by the GTech human ID at a distance project. Table 3.2 gives the description of all the five probe sets. Figure 3.9 shows some example video frames from the GTech database. Figure 3.10 shows some example silhouettes from the GTech database.

**GC Database:** The HumanID gait challenge database has been designed by the University of South Florida to evaluate the performance of gait recognition systems [69]. The database consists of data sets recorded under varying conditions so that the effects
Chapter 3. Determinants in Gait Recognition

Table 3.1: Summary of gait databases

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Subj.</th>
<th>Seq.</th>
<th>Scene</th>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSD</td>
<td>6</td>
<td>42</td>
<td>Wall Background</td>
<td>Time</td>
</tr>
<tr>
<td>CMU MoBo</td>
<td>25</td>
<td>600</td>
<td>Indoor, Treadmill</td>
<td>Viewpoint, Speed, Surface incline, Carrying condition</td>
</tr>
<tr>
<td>GTech</td>
<td>15</td>
<td>268</td>
<td>Indoor, Outdoor</td>
<td>Time, Viewpoint</td>
</tr>
<tr>
<td>Maryland</td>
<td>55</td>
<td>222</td>
<td>Outdoor, Top mounted</td>
<td>Time, Viewpoint</td>
</tr>
<tr>
<td>MIT</td>
<td>24</td>
<td>194</td>
<td>Indoor</td>
<td>Time</td>
</tr>
<tr>
<td>SOTON</td>
<td>28</td>
<td>112</td>
<td>Indoor, Background</td>
<td>Time</td>
</tr>
<tr>
<td>Gait Challenge</td>
<td>122</td>
<td>1870</td>
<td>Outdoor</td>
<td>Time, Viewpoint, Surface, Shoe, Carrying condition</td>
</tr>
</tbody>
</table>

of different environmental factors on performance can be evaluated. This database also provides a common ground to compare the performance of different algorithms under the same conditions. The most recent version of the database Nov-2001 consists of sequences of 122 subjects. The data was recorded using two digital miniDV video cameras when the subjects walked in an elliptical course. The data was later transcoded from DV to 24-bit RGB with 720 x 480 PPM file per frame. The length of the videos is about 200 frames. Figure 3.11 shows some example video frames from the GC database. The total size of the video gait database is about 1.2 Tera bytes.

The challenge problem consists of a set of 12 experiments to investigate the effect of five factors on performance. These five factors are studied both individually and in different combinations. A Baseline algorithm is also given which is based on the correlation between the probe and gallery silhouette sequences. Table 3.3 shows the number of gait sequences with different covariate combinations and the last column gives
Table 3.2: GTech probe sets

<table>
<thead>
<tr>
<th>Probe Set</th>
<th>Recording Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe 1</td>
<td>All sequences except gallery sequences</td>
</tr>
<tr>
<td>Probe 2</td>
<td>Indoor angle 0</td>
</tr>
<tr>
<td>Probe 3</td>
<td>Indoor angle side near</td>
</tr>
<tr>
<td>Probe 4</td>
<td>Indoor angle side far</td>
</tr>
<tr>
<td>Probe 5</td>
<td>Outdoor angle</td>
</tr>
</tbody>
</table>

The number of subjects that were common in both May and November data collection. The number of probe set for each of the gait challenge experiments are given in Table 3.4 for both May and November data sets. The symbols used in the tables are explained as follows:

**Surface type:** G for grass and C for concrete

**Camera:** R for right and L for left

**Shoe type:** A or B

**Briefcase:** NB for no briefcase and BF for carrying a briefcase

**Acquisition Time:** M for May and N for November

The May 2001 version of the database is used in our experimental analysis. Table 3.5 shows the details of gallery and probes for experiments A–G. Figure 3.12 shows some example silhouettes from the GC database.

### 3.3.2 Feature Vector Normalization

The dynamic range of each component in the feature vector is different because each represents a different physical quantity. The feature vector components having higher
values will overshadow the components with lower values in similarity calculation. Therefore the features are normalized before the application of the similarity measure. The normalization process ensures that each component of the feature has equal emphasizes.

Let $V$ be the sequence of values to be normalized. Then the sequence can be normalized to $[0, 1]$ range by applying Equation 3.9 [70]:

$$V_i = \frac{V_i - V_{\text{min}}}{V_{\text{max}} - V_{\text{min}}} \quad (3.9)$$

Where $V_{\text{min}}$ and $V_{\text{max}}$ are the minimum and the maximum values of the sequence and $V_i$ represents an element of the sequence. This normalization procedure is very simple but it does not provide desirable results. For example, let us consider a sequence of values $[1.4, 1.8, 2.7, 2.3, 200]$. By using the above normalization, most of the $[0, 1]$ range will be taken by a single quantity of 200. The other values $[1.4, 1.8, 2.7, 2.3]$ will be wrapped.
Figure 3.10: Example silhouettes from GTech database

Figure 3.11: Example video frames from GC database (a,b) Concrete surface (b,c) Grass surface

with a very small range. A better way to normalize the sequence is to consider it being generated by a Gaussian distribution. In this procedure, we calculate the mean $\mu$ and standard deviation $\sigma$ of the sequence. The sequence is then normalized by Equation 3.10 [70]:

$$V_i = \frac{V_i - \mu}{\sigma}$$  \hspace{1cm} (3.10)

This procedure maps most of the values of the feature sequence $V$ in $[-1, 1]$ range. The advantage of this normalization is that a few abnormal values occurring in the sequence will not bias the importance of other values. We use this normalization technique to normalize the feature vectors.
Table 3.3: Number of sequences for each possible combination

<table>
<thead>
<tr>
<th>Surface</th>
<th>Carry</th>
<th>Shoe</th>
<th>Camera</th>
<th>Time</th>
<th>Common</th>
</tr>
</thead>
<tbody>
<tr>
<td>M or N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concrete</td>
<td>NB</td>
<td>A</td>
<td>(L,R)</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>B</td>
<td>(L,R)</td>
<td>60</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>A</td>
<td>(L,R)</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>B</td>
<td>(L,R)</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Grass</td>
<td>NB</td>
<td>A</td>
<td>(L,R)</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>B</td>
<td>(L,R)</td>
<td>54</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>A</td>
<td>(L,R)</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BF</td>
<td>B</td>
<td>(L,R)</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.12: Example silhouettes from GC database

3.3.3 Feature Matching

We use nearest neighborhood approach for classification of gait. The similarity criterion is very critical in this approach. The metric distance between the feature vectors of the probe and the gallery silhouette sequences is commonly used for similarity measurement. The Minkowski-form distance is defined based on the $L_p$ norm [71]:

$$L_p(p, g) = (\sum_{i=0}^{N-1} |p_i - g_i|^p)^{\frac{1}{p}} = \|p - g\|_p$$

(3.11)

where $p$ and $g$ are vectors of dimension $N$ and $p_i$ and $g_i$ are their $i$th components respectively. The above equation is the general form of the distance metric. If $p = 1,$
Table 3.4: Probe set for each of the gait challenge experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Probe</th>
<th># of</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(G, A, L, NB, M/N)</td>
<td>122</td>
<td>View</td>
</tr>
<tr>
<td>B</td>
<td>(G, B, R, NB, M/N)</td>
<td>54</td>
<td>Shoe</td>
</tr>
<tr>
<td>C</td>
<td>(G, B, L, NB, M/N)</td>
<td>54</td>
<td>Shoes, View</td>
</tr>
<tr>
<td>D</td>
<td>(C, A, R, NB, M/N)</td>
<td>121</td>
<td>Surface</td>
</tr>
<tr>
<td>E</td>
<td>(G, B, R, NB, M/N)</td>
<td>60</td>
<td>Surface, Shoe</td>
</tr>
<tr>
<td>F</td>
<td>(C, A, L, NB, M/N)</td>
<td>121</td>
<td>Surface, view</td>
</tr>
<tr>
<td>G</td>
<td>(C, B, L, NB, M/N)</td>
<td>60</td>
<td>Surface, Shoe, View</td>
</tr>
<tr>
<td>H</td>
<td>(G, A, R, BF, M/N)</td>
<td>120</td>
<td>Briefcase</td>
</tr>
<tr>
<td>I</td>
<td>(G, B, R, BF, M/N)</td>
<td>60</td>
<td>Shoe, Briefcase</td>
</tr>
<tr>
<td>J</td>
<td>(G, A, L, BF, M/N)</td>
<td>120</td>
<td>View Briefcase</td>
</tr>
<tr>
<td>K</td>
<td>(G, A/B, R, NB, N)</td>
<td>33</td>
<td>Time, Shoe, Clothing</td>
</tr>
<tr>
<td>L</td>
<td>(C, A/B, R, NB, N)</td>
<td>33</td>
<td>Surface, Time, Shoe, Clothing</td>
</tr>
</tbody>
</table>

then the distance is known as the **City-block** or **Manhattan** distance defined as follows:

\[
L_1(\mathbf{p}, \mathbf{g}) = \sum_{i=0}^{N-1} |p_i - g_i| = \|\mathbf{p} - \mathbf{g}\|_1
\]  
(3.12)

Another famous distance metric is **Euclidean** or \(L_2\) norm defined when \(p = 2\):

\[
L_2(\mathbf{p}, \mathbf{g}) = \left( \sum_{i=0}^{N-1} |p_i - g_i|^2 \right)^{\frac{1}{2}} = \|\mathbf{p} - \mathbf{g}\|_2
\]  
(3.13)

**Euclidean** and **City-block** distance measure only the difference in the lengths of the two vectors. In some cases, the angle between the vectors may be more significant for purpose of similarity. The cosine distance measures the difference in the direction of two vectors irrespective of their length. The **cosine distance** is defined in Equation 3.14 [71]:

\[
d_{\cos}(\mathbf{p}, \mathbf{g}) = \frac{\mathbf{p}^T \mathbf{g}}{\|\mathbf{p}\|_2 \cdot \|\mathbf{g}\|_2}
\]  
(3.14)
Table 3.5: Probe set for gait challenge experiments A–G

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Probe</th>
<th># of</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gallery</td>
<td>(G, A, R)</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>(G, A, L)</td>
<td>71</td>
<td>View</td>
</tr>
<tr>
<td>B</td>
<td>(G, B, R)</td>
<td>41</td>
<td>Shoe</td>
</tr>
<tr>
<td>C</td>
<td>(G, B, L)</td>
<td>41</td>
<td>Shoes, View</td>
</tr>
<tr>
<td>D</td>
<td>(C, A, R)</td>
<td>70</td>
<td>Surface</td>
</tr>
<tr>
<td>E</td>
<td>(G, B, R)</td>
<td>44</td>
<td>Surface, Shoe</td>
</tr>
<tr>
<td>F</td>
<td>(C, A, L)</td>
<td>70</td>
<td>Surface, view</td>
</tr>
<tr>
<td>G</td>
<td>(C, B, L)</td>
<td>44</td>
<td>Surface, Shoe, View</td>
</tr>
</tbody>
</table>

Where $p^T$ is the transpose of vector $p$. We can see that this is very similar to the correlation coefficient. In the following experimental evaluation, we have use the city-block distance measure. The selection is based on the empirical analysis.

3.3.4 Comparative Performance of Dynamic Features

Silhouette sequences are processed frame by frame for the extraction of dynamic gait features. The silhouette frames are first processed by median filtering to reduce outliers. We also estimate the gait period from the autocorrelation function of the silhouette area signal. Speed is normalized by ensuring the same number of frames in each gait cycle for all silhouette sequences. The six area signals are extracted and EMD algorithm is applied to reduce the noise. DCT coefficients of autocorrelation functions of six reconstructed area signals are calculated. We use first 35 DCT coefficients as gait features for the GC database while only first 6 coefficients are used for GTech database. The number of DCT coefficients were chosen empirically to achieve the best recognition performance. The feature vectors are normalized using the procedure given in the Section 3.3.2. CMS values are used to evaluate the performance of different dynamic gait features. Each of
the probe sequence features are compared with the features of gallery sequences. The gallery sequence set is sorted according to the similarity to the probe sequences. We use nearest neighborhood approach in combination with city block distance measure.

Table 3.6 presents the performance evaluation of features extracted from the six area signals at rank 1 and rank 5 for GC database. We use experiment A to analyze the recognition potential of dynamic features extracted from different parts of the silhouettes. Experiment A is chosen due to the following reasons:

- Both Gallery and Probe set A contain all the 71 subjects
- They are recorded under similar conditions except a different viewpoint. This eliminates the effect of other covariates which can skew the results.

It is observed from these results that best performance of 97.18% is achieved from the features extracted from \( a_6 \). This area signal represents the dynamics of lower leg during gait motion. The second most significant results of 78.87% are achieved from \( a_3 \) which represents the lower arm dynamics. We achieve a recognition performance of 73.24% for both \( a_4 \) and \( a_5 \) features. Similarly the rank 1 recognition performance of \( a_1 \) and \( a_2 \) features is also same at 53.52%. The rank 1 results indicate that thigh movement and knee movement are of equal importance in gait recognition. However, the performance of thigh and knee features is slightly lower than the features extracted from the lower arm dynamics.

Similar result pattern is obtained at rank 5 as shown in Table 3.6. In case of rank 5, \( a_6 \) and \( a_3 \) features provide the best recognition performance of 100%. The recognition rate of \( a_4 \) and \( a_5 \) features is slightly lower at 98.59%. The lowest performance of 92.96% is obtained from \( a_2 \) features. The recognition performance of \( a_1 \) features is slightly higher than \( a_2 \) feature set at 94.37%. Figure 3.13 shows the CMC plot for the DCT features of six area. The recognition performance of the dynamic features extracted from \( a_3 \) and \( a_6 \) is superior to the other feature sets. These results partially support the traditional notion
Table 3.6: Comparison of features at rank 1 and rank 5 for GC database

<table>
<thead>
<tr>
<th>Area Signal</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>53.52</td>
<td>94.37</td>
</tr>
<tr>
<td>$a_2$</td>
<td>53.52</td>
<td>92.96</td>
</tr>
<tr>
<td>$a_3$</td>
<td>78.87</td>
<td>100</td>
</tr>
<tr>
<td>$a_4$</td>
<td>73.24</td>
<td>98.59</td>
</tr>
<tr>
<td>$a_5$</td>
<td>73.24</td>
<td>98.59</td>
</tr>
<tr>
<td>$a_6$</td>
<td>97.18</td>
<td>100</td>
</tr>
</tbody>
</table>

of significance of leg dynamics in gait recognition. It is also observed that the dynamics of the lower arm is very important in determining the gait pattern of the human subjects in an appearance based set up.

The results for the GTech database are given in Table 3.7. We use Probe 2 to evaluate the features from six area signals due to the reason that both Gallery and Probe 2 sequences were recorded indoor. The camera angle for both Gallery and Probe 2 sequences is also the same. As mentioned earlier, this enables us to watch the recognition performance of dynamic features extracted from different parts of the silhouettes without the bias of other covariates. The recognition results for GTech database vary from GC database and the best performance is achieved for features extracted from $a_1$. The performance of features extracted from leg motion is the lowest. The features extracted from $a_1$ and $a_2$ perform almost equivalently giving a recognition rate of 47.73% and 46.6% respectively at rank 1. Similarly the recognition rate for $a_1$ and $a_2$ features at rank 5 is 84.1% and 81.2% respectively. The rank 1 performance of $a_6$ features is slightly higher than features extracted from $a_5$. However, the rank 5 performance of features extracted from $a_6$ is 10.23% higher than those extracted from $a_5$. These results indicate that the dynamic features extracted from upper arm and shoulder part of the silhouettes
Figure 3.13: CMC of six area based features for GC database

posses the most distinguishing power for GTech database. The reason for the difference in observation between GC and GTech database lies in the nature of silhouettes. The quality of silhouettes is better in case of GC database. The silhouettes from GTech database has more holes and missing parts in the lower leg and around knee points. This missing information degrades the quality of features resulting in lower recognition performance of dynamic features extracted from lower half of the silhouettes. The CMC plot shown in 3.14 also establishes the superiority of features extracted from $a_1$ and $a_2$.

### 3.4 Summary

Human locomotion is a complex phenomenon involving the coordination of different limbs as the body translates from one point to another. The static configuration of the body such as the widths and lengths of different limbs have been shown of great importance in
determining the gait pattern of the individuals. The contribution of dynamics of different parts of human body has not been studied explicitly in an appearance based recognition set up. We analyzed the recognition performance of dynamic features from different parts of the body. It is shown by experimental evaluation that dynamics of lower leg and lower arm are of utmost importance for building an efficient gait recognition system. However, in case of GTech database the quality of features extracted from leg portion of silhouettes performed lower than those extracted from upper arm and shoulders. The motion of lower half of the body has always been considered more important in the determination of gait pattern. However, we have found that lower arm movement also plays an important role in gait recognition. We also showed evidence from our results that the dynamics of upper arm and shoulders are also very useful for extraction of quality features for gait recognition. We will use these results in the next chapter to extract new dynamic gait features and combine them with some existing features to improve the recognition rate.
Figure 3.14: CMC of six area based features for GTech database
Chapter 4

Correlation based Gait Recognition

4.1 Introduction

The potential of correlation based features for gait recognition was established in Chapter 3. The technique will be further explored and a comprehensive experimental evaluation will be presented in this chapter. The features will be extracted from both 2nd. order and 3rd. order autocorrelation functions of the area signals. The correlation based features are also combined with Fourier descriptor features to improve the recognition performance. The performance of the purposed features will be compared with some representative methods reported in literature using the silhouette sequences from the GTech database described in detail in the preceding chapter.

4.2 Silhouette Preprocessing

The quality of the silhouettes is very critical in our method because it is based on the appearance. However, it is noted that the silhouettes are quite noisy. We noticed outliers, broken legs and noisy edges. Therefore, in order to capture reliable gait motion dynamics, we need to improve the quality of the silhouettes. The preprocessing is performed in two steps. In the first step, median filtering is applied to get rid of the outliers from the
silhouettes. The second step involves the application of morphological image processing to smooth the edges and connect the broken legs. Both pre-processing steps are explained below in detail.

We use median filtering with a mask of size 5x5 to filter the silhouettes to get rid of the outliers. The output from the median filter is binarized by simple thresholding to obtain smoothed silhouettes. Figure 4.1 shows three example silhouettes from the GTech database containing such outliers and Figure 4.2 shows the same silhouettes after median filtering. After median filtering, we apply morphological closing operation to fuse the narrow gaps and smooth the contours of the silhouettes. The morphological closing operation is defined as follows [67]:

$$\hat{B} = B \bullet S = (B \oplus S) \ominus S$$  \hspace{1cm} (4.1)

Where $B$ is the binary silhouette before the closing operation and $\hat{B}$ is the binary silhouette after the closing operation. $S$ is the matrix of 0s and 1s and is called the structuring element. $\oplus$ and $\ominus$ represent morphological dilation and erosion operations respectively.
The dilation and erosion operations are defined as follows.

\[(B \oplus S) = \{z|[(\hat{S})_z \cap B] \subseteq B\}\] (4.2)

\[(B \ominus S) = \{z|(S)_z \subseteq B\}\] (4.3)

\[\hat{S} = \{w|w = -s, \text{ for } s \in S\}\] (4.4)

Where \(\cap\) and \(\subseteq\) mean intersection and subset respectively and \(z\) is an element of \(Z^2\) or 2D integer space. \(\hat{S}\) is the reflection set of \(S\). Figure 4.3 shows the results of closing operation on one of the silhouettes from the database. Although the operation removes the lower part of the legs yet this removal does not affect the gait feature extraction process. The reason is that we use two area masks of width 15 pixels each on the upper leg and shoulder portion to extract the gait dynamics. The area mask placement is decided based on the results of the comparative analysis given in Chapter 3. The feature extraction process is described in detail in the next section.

(a) Before closing \hspace{1cm} (b) After closing

Figure 4.3: Morphological closing on a silhouette

4.3 Feature Extraction

Silhouette sequences are processed frame by frame for the extraction of gait features. The video sequences in the GTech database were recorded with a fixed camera. The distance
between the camera and the subjects varies during the walking process. Therefore, the position and height of the silhouettes in the frames also varies in the silhouette sequence. Silhouettes are centred in each frame of the silhouette sequence and also scaled to the same height preserving the aspect ratio.

The walking speed of the subjects is also different in different silhouette sequences. Therefore gait period is not constant across gait sequences. This change in gait period will effect the calculation of correlation based features because of their dependency on the frequency of area signals. Speed normalization is performed to neutralize the effect of the walking speed on gait features. This is achieved by ensuring the same number of frames in each gait cycle for all silhouette sequences. The estimation of gait period is required for speed normalization. The noise in the area signal makes the calculation of the gait period ambiguous. We estimate the gait period from the autocorrelation function of the silhouette area signal. The gait period is given by $2T$, where $T$ is the smallest time lag other than $l = 0$, corresponding to the local maxima of autocorrelation function $R_a[l]$. Figures 4.4 and 4.5 show a typical silhouette area function and its corresponding autocorrelation function respectively. It is clear from these figures that the calculation of gait period is easier from $R_a[l]$ compared to the original noisy area signal [40].

### 4.3.1 Correlation Features

The upper leg and shoulder motion is captured by placing two area masks of width 15 pixels on the corresponding parts of the binary silhouettes. The number of pixels of the binary silhouettes under these two masks is calculated. The process is repeated for each binary silhouette in the gait sequence and we obtain two area signals of length $N$, where $N$ is the number of frames in the gait sequence. Let

$$b[i, j] = \begin{cases} 
1 & \text{if pixel } [i, j] \text{ belongs to the foreground} \\
0 & \text{Otherwise} 
\end{cases} \quad (4.5)$$
Chapter 4. Correlation based Gait Recognition

Figure 4.4: Silhouette area signal

\[
m_1[i,j] = \begin{cases} 
1 & \text{if } J_1 \leq j < J_1 + 15 \\
0 & \text{Otherwise}
\end{cases} \quad (4.6)
\]

\[
m_2[i,j] = \begin{cases} 
1 & \text{if } J_2 \leq j < J_2 + 15 \\
0 & \text{Otherwise}
\end{cases} \quad (4.7)
\]

\(m_1\) and \(m_2\) are the two area masks and \(J_1\) and \(J_2\) are the starting Y-coordinate values for the two area masks. \(J_1\) and \(J_2\) are chosen in such a way that the masks are placed on the upper leg and shoulder potion of the silhouettes. Now the two area functions \(a_1\) and \(a_2\) are given as:

\[
a_1[n] = \sum_{i,j} b_n[i,j]m_1[i,j] \quad (4.8)
\]

\[
a_2[n] = \sum_{i,j} b_n[i,j]m_2[i,j] \quad (4.9)
\]

where \(n\) is the frame number.

The next step is to calculate both 2nd. and 3rd. order autocorrelation functions of
the two area signals. We take a 1-D diagonal slice of the 3rd. order autocorrelation function. The 2nd. order autocorrelation, $R_{ap}[l]$, is calculated by Equation 3.6 given in Section 3.2. 1-D diagonal slice of the 3rd. order autocorrelation, $\hat{R}_{ap}[l]$, of area signals is calculated as follows:

$$\hat{R}_{ap}[l] = \sum_{n} a(n)a^2[n+l]$$  \hspace{1cm} (4.10)$$

The gait signatures are then derived by taking the DCT, $T_p[k]$, of the auto-correlation functions defined by Equations 3.7 and 3.8 in Section 3.2. The mean and standard deviation of the two area functions are also used as gait features.

### 4.3.2 Fourier Descriptor Features

The shape of the silhouette changes during the gait cycle. The shape changes during the gait cycle can be calculated using the boundary descriptor such as FD. FD is invariant to scale, rotation and translation and is robust in the presence of noise. FD based features were successfully used for gait recognition purposes [39].
The boundary of a silhouette can be represented digitally by \( N \) points, \((x[n], y[n])\), for \( n = \{0, 1, \cdots, N - 1\} \), where \( N \) is the length of the boundary. These boundary points can be represented in terms of complex numbers by considering the image as a complex plane. The \( X \)-axis represents the real axis while \( Y \)-axis is taken as the imaginary axis. Complex coordinate function \( z[n] = x[n] + jy[n] \) expresses the boundary points as a one dimensional signal.

In order to calculate the FD based gait signature, we first find the centroid \((x_c, y_c)\) of each silhouette in the gait sequence. Centroid Contour Distance (CCD) is then determined as follows:

\[
r[n] = \sqrt{(x[n] - x_c)^2 + (y[n] - y_c)^2}
\] (4.11)

We take the Fourier transform of the CCD \((r[n])\) as defined by Equation 4.12. The Discrete Fourier Transform (DFT) of \( r[n] \) is calculated as given below:

\[
r[k] = \frac{1}{N} \sum_{n=0}^{N-1} r[n] e^{-j2\pi kn/N}
\] (4.12)

for \( k = \{0, 1, \cdots, K\} \). The Fourier coefficients are closely related to FD [15]. Global shape is captured by the first few low frequency terms in modified FDs, while higher frequency FDs determine the details of the shape. The advantage of using CCD is the rotation and translation invariance. The CCD Fourier spectra are more concentrated around the origin of the polar space. This is also a desirable characteristic because a few FDs can describe the shape efficiently. However, the FD of the CCD is not scale invariant which is required for the gait signature. The scale invariance is achieved by dividing all the FDs by DC component. The DC component is ignored and is not used as a gait feature because we are interested in the dynamics of shape change.

The changes in the shape of the silhouettes during the gait cycle are imbedded in the FDs at different frequencies. We take the temporal mean and standard deviation of the FDs of different frequencies and use them as a gait signature. The boundary of each silhouette in the gait sequence is represented by the same number of points to get the
Chapter 4. Correlation based Gait Recognition

FDs at the same frequencies.

4.4 Experimental Evaluation

This section presents the results of experimental evaluation using the GTech database described in detail in Section 3.3.1. We conduct experiments using different combinations of 3 types of gait features to empirically select the best combination in terms of recognition performance. The performance of this combined gait feature set is then compared with other state of the art techniques. The five probe sets used in this experimental evaluation are listed in Table 3.2.

4.4.1 Feature Database

The autocorrelation of the distance function is estimated for the positive lags only. The 2nd. order autocorrelation is calculated for 40 lags while the 1-D diagonal slice of the 3rd. order autocorrelation is calculated for the first 20 lags. This is due to the short length of the area signal and the estimates will be noisy for higher lags. Then we take 64-point DCT of the 4 autocorrelation and 2 cross-correlation functions to calculate the Correlation Features (CF). The mean and standard deviation of the area signals form the Distance Feature (DF) vector of dimension 4.

The boundary of the silhouettes is calculated by using the Canny edge detection algorithm [73]. The Canny edge detector uses Gaussian convolution to smooth the image in the first step to maximize the signal to noise ratio. It achieves good localization to accurately mark edges with 2-D first derivative operator. Edges give rise to ridges in the gradient magnitude image. The algorithm then minimizes the number of responses to a single edge by tracking along these regions and suppressing any pixel that is non-maximum. The gradient array is reduced by hysteresis. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is made a non-edge by setting it...
to zero. If the magnitude is above the higher threshold, it is made an edge. And if the magnitude is in between the 2 thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above higher threshold. The image is then converted to a binary image.

For every silhouette in the binary gait sequence, the boundary is represented in terms of CCD and then 256 point DFT of the CCD is taken. Then we choose the first 129 FDs because DFT of a real sequence is symmetric. We then normalize them by dividing with the DC component. The DC component is normalized by dividing it by the area of the silhouette. We take the mean and standard deviation of the FDs at corresponding frequencies to use them as gait features.

The 3 types of gait features have different dynamic ranges and are normalized as described in section Section 3.3.2. The classification is performed using the city block distance defined in Equation 3.12 using nearest neighbourhood approach.

4.4.2 Results and Discussion

Table 4.1 summarizes the recognition results obtained by using different combinations of CFs, FDs and DFs for probe 1. Probe 1 consists of all the silhouette sequences in the GTech database except the gallery sequences. It is observed that the performance of 2nd. order correlation features outperforms that of FD features by 2.98% at rank 1 and by 12.69% at rank 5. These results indicate that correlation of the upper leg and shoulder dynamics is an important new gait signature which can be combined with other signatures to achieve better recognition results. It is further noted that 3rd. order correlation of area signals does not perform well and its performance is the lowest compared to the performance of FD and 2nd. order CF. The performance of 3rd. order correlation features is 19.94% lower at rank 1 and 32.83% lower at rank 5 compared to that of 2nd. order CF. When we combine the 2nd order CF and FD features with the two area statistics, a CMS value of 66.05% and 82.84% is obtained at rank 1 and rank 5 respectively. We select
Table 4.1: Recognition results for probe 1

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd. order CF + DF</td>
<td>77</td>
<td>52.61</td>
<td>84.33</td>
</tr>
<tr>
<td>3rd. order CF + DF</td>
<td>77</td>
<td>33.21</td>
<td>51.5</td>
</tr>
<tr>
<td>FD + DF</td>
<td>38</td>
<td>49.63</td>
<td>71.64</td>
</tr>
<tr>
<td>2nd. order CF + DF + FD</td>
<td>115</td>
<td>66.05</td>
<td>82.84</td>
</tr>
<tr>
<td>Baseline [32]</td>
<td>silhouettes</td>
<td>24.25</td>
<td>50.37</td>
</tr>
<tr>
<td>Mass Vector [72]</td>
<td>128 x N</td>
<td>73.51</td>
<td>92.16</td>
</tr>
</tbody>
</table>

this combination of features for further experiments and comparison with other works reported in literature. Figure 3 shows the CMC plot for the 4 features combinations.

The Baseline algorithm has been used as a benchmark for performance comparison in the recent works in gait recognition. The performance of our method is much higher than the Baseline algorithm. The Baseline algorithm performs very poorly on the GTech database. The performance of the proposed feature set is 41.4% higher at rank 1 and 32.47% higher at rank 5 than that of Baseline algorithm. However, the mass vector approach reported in [72] performs slightly better than our method at both rank 1 and rank 5. The dimension of the gait feature vector is lower in our technique which is 115 compared to 128 × N for the mass vector method. Here N represents the number of frames in the silhouette sequence.

We summarize the values of CMS for probe 3 in Table 4.2. The results indicate that CF method achieves great improvement in recognition performance over Baseline algorithm. The performance is 55.56% and 51.11% higher than the Baseline at rank 1 and rank 5 respectively. Our method also has much lower computational cost at recognition stage than that of Baseline algorithm which uses the whole silhouettes as features. The performance of mass vector method surpasses our method both for rank 1 and rank 5.
Chapter 4. Correlation based Gait Recognition

The rank 1 performance of GTech method is lower than our CF method by 27.78%. But at rank 5, GTech method also performs better than the proposed method. Figure 4.7 shows the CMC curve for probe 3.

Rank 1 and rank 5 results for probe 4 are presented in Table 4.3. It is observed that in this category GTech features perform better than CF by 8.37% at rank 1. However the combined feature set of CF and FD perform better than the GTech features by 5% at rank 1. The performance of our method is much higher than the Baseline benchmark at both rank 1 and rank 5. But the mass vector technique outperforms our correlation based method both at rank 1 and rank 5. We also note an interesting result from this probe, the correlation based features perform higher at rank 5 compared to the combined CF and FD feature set. The CMC curve showing the recognition performance for probe 4 is shown in Figure 4.8.

In the end, we perform recognition experiments using probe 5. The results are pro-
Table 4.2: Recognition results for probe 3

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd. order CF + DF + FD</td>
<td>115</td>
<td>77.78</td>
<td>93.33</td>
</tr>
<tr>
<td>Baseline silhouettes</td>
<td></td>
<td>22.22</td>
<td>42.22</td>
</tr>
<tr>
<td>Mass Vector</td>
<td>128 x N</td>
<td>97.78</td>
<td>100</td>
</tr>
<tr>
<td>GTech method [68]</td>
<td>4</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4.3: Recognition results for probe 4

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd. order CF + DF</td>
<td>77</td>
<td>66.67</td>
<td>93.33</td>
</tr>
<tr>
<td>2nd. order CF + DF + FD</td>
<td>115</td>
<td>80</td>
<td>80.89</td>
</tr>
<tr>
<td>Baseline silhouettes</td>
<td></td>
<td>11.11</td>
<td>46.67</td>
</tr>
<tr>
<td>Mass Vector</td>
<td>128 x N</td>
<td>93.33</td>
<td>100</td>
</tr>
<tr>
<td>GTech method</td>
<td>4</td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

vided in Table 4.4. Clearly the results for this probe are the lowest across all techniques. This is expected because the gallery sequences were recorded indoor while the probe 5 sequences were recorded outdoor. Outdoor sequences suffer from a variety of changing conditions such as background and lighting. The performance of CF method is higher than the Baseline benchmark by 13.34% and 15.55% at rank 1 and rank 5 respectively. The recognition rate is also comparable with the other two techniques. The mass vector technique performs slightly lower than our method at rank 1 for this probe. On the other hand the rank 5 performance of mass vector method is consistently higher than our method for all probes. GTech method performs really well in this case. This is due to the reason that GTech method is model based and hence is less sensitive to the varying
conditions compared to the other appearance based methods. The CMC curve plotted in Figure 4.9 shows the recognition results for probe 5 at different ranks.

4.5 Summary

In this chapter, we presented a new method for the extraction of gait features from silhouette sequences. The method is based on the correlation analysis of upper leg and shoulder motion. Experimental evaluations indicate that the 2nd order correlation based gait signatures possess high discrimination power for gait recognition. We also presented a comprehensive comparison of performance between different combinations of gait features. The comparative study shows that 2nd order correlation features are superior not only to the 3rd order correlation (1-D diagonal slice) based features but also outperform FD. Our method achieved much better performance than the Baseline algorithm bench-
mark. GTech method showed much better performance in case of probe 5 or outdoor versus indoor. This leads us to conclude that our proposed CF method is more sensitive to camera angle and recording conditions similar to other appearance based techniques.

It is further observed that the major drawback of correlation based gait features is their high sensitivity to the quality of silhouettes. This sensitivity is inherent to all appearance based approaches as is evident from the results. However, better preprocessing techniques can be used to overcome this problem. We use morphological closing operation to smooth the silhouettes but the problem of finding the optimum structuring element is not addressed in this work. It is also noted that by implementing a simple square structuring element with 9 neighbors results in over-smoothing of the silhouettes in some cases. Another important aspect observed from the above experiments is the dependency of recognition results on the database. The Baseline algorithm does not perform very well on GTech database and its recognition results drop significantly compared to GC.
Chapter 4. Correlation Based Gait Recognition

Table 4.4: Recognition results for probe 5

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd. order CF + DF + FD</td>
<td>115</td>
<td>27.78</td>
<td>63.33</td>
</tr>
<tr>
<td>Baseline silhouettes</td>
<td>14.44</td>
<td>47.78</td>
<td></td>
</tr>
<tr>
<td>Mass Vector</td>
<td>128 x N</td>
<td>25.56</td>
<td>76.67</td>
</tr>
<tr>
<td>GTech method</td>
<td>4</td>
<td>38</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 4.9: CMC for probe 5

database.
Chapter 5

Wavelet Analysis of Human Gait

5.1 Introduction

This chapter will investigate the potential of using wavelet transform to decompose the area signals and extract features for gait recognition. Studies have shown that human gait is quasi-periodic and there are slight changes in the fundamental frequency and amplitude over time [74]. It is observed from Figure 4.4 that the frequency content of the area signal varies with time. The direct application of Fourier analysis technique to such signals may lead to incomplete information extraction. This motivates us to use time-frequency analysis to analyze these signals. We use Discrete Wavelet Transform (DWT) to perform such analysis. Short Time Fourier Transform (STFT) can also be used to perform time-frequency analysis but it uses a fixed size rectangular tiling of the time-frequency plane. In contrast, the wavelet transform uses rectangles of variable dimensions but constant area. Both STFT and wavelet transform use finite duration basis functions to estimate the transform. But in case of wavelet transform the length of the support of basis function is a function of frequency. We use wavelet transform because of this desirable property. Figure 5.1 shows the sampling grid for wavelet transform which is dyadic in nature. In comparison, the sampling grid in case of STFT is uniform as
Chapter 5. Wavelet Analysis of Human Gait

illustrated in Figure 5.2.

Figure 5.1: Sampling grid for time-scale plane

Figure 5.2: STFT coverage of Time-Frequency plane

Appendix C discusses the basics of Continuous Wavelet Transform (CWT) essential to understand DWT. DWT is also concisely described in Appendix C without going into the rigorous mathematical details which is beyond the scope of this dissertation.
5.2 Extraction of Wavelet Features

The binary silhouette sequences are processed frame by frame. For each frame in the binary gait sequence, we first perform the median filtering to reduce the outliers. After the median filtering, the area of the lower half of the silhouette is calculated by counting the number of pixels in the lower half of the binary silhouette. This gives us an area signal of length equal to the number of frames in the sequence. We then decompose this area signal into wavelet sub-bands by applying different 1-D Daubechies (db) wavelet kernels. A wavelet filter dbM (where M is a positive integer) has compact support and length 2M. This means that by increasing M, we are increasing the length of the filter. This translates to lower spatial resolution with increasing M. Furthermore as we increase the order of the db filter, its bandwidth is also increased and it allows finer details to be analyzed. Therefore there is a tradeoff between the two factors, spatial resolution and bandwidth and db4 performed superior to other wavelets in our empirical analysis.

A d level wavelet decomposition gives us d detailed frequency sub-bands and one low frequency sub-band. We calculate mean, standard deviation, skewness and kurtosis for each sub-band. These statistics are defined as under:

\[
\mu = \frac{1}{N} \sum_{i=1}^{N} w_i \quad (5.1)
\]

\[
\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (w_i - \mu)^2} \quad (5.2)
\]

\[
\text{skewness} = \frac{1}{(N-1)\sigma^3} \sum_{i=1}^{N} (w_i - \mu)^3 \quad (5.3)
\]

\[
\text{kurtosis} = \frac{1}{(N-1)\sigma^4} \sum_{i=1}^{N} (w_i - \mu)^4 \quad (5.4)
\]

Where \(w_i\) are the wavelet coefficients, \(\mu\) is the mean and \(\sigma\) stands for standard deviation. \(N\) represents the number of wavelet coefficients. We perform 3-level wavelet decomposition and hence the gait signature is 16 dimensional. The proposed gait signature extraction technique can be summarized in 5 steps as shown in Algorithm 1 below.
Algorithm 1: Feature extraction using DWT

1. Median filtering for noise reduction.

2. Determination of the area of lower half of each silhouette in the sequence by calculating the number of non zero pixels.

3. Re-sampling the area signals to match the length of the gallery sequences.

4. Application of DWT to area signals (3-level decomposition)

5. Calculation of mean, standard deviation, skewness and kurtosis of each wavelet subband.

5.3 Experimental Evaluation

Experimental evaluation is performed using GTech database explained in Section 3.3.1. The probe sets for this experimental assessment are given in Table 3.2. We use the standard CMS values to measure the efficacy of the proposed feature set and its comparison with Baseline, mass vector and GTech methods. DWT features are combined with FD features and nearest neighbourhood approach is used for classification using Euclidean distance as a similarity function. Euclidean distance was defined in Section 3.3.3 by Equation 3.13.
5.3.1 Performance of Different Wavelet Kernels

We first compare the identification performance of 3 wavelet kernels both at rank 1 and rank 5. Table 5.1 shows the results for db1, db2 and db4 based features at rank 1. Here Normalized stands for speed normalization and Raw means features are extracted without any speed normalization. The speed is normalized by the following two steps.

- Gait period of each sequence is determined from the autocorrelation function of the area signal as described in Section 4.3.

- Features are extracted from only one gait period of the area signal after re-sampling it to a common length.

It is noted from the Table 5.1 that the results do not differ significantly for the normalized and non-normalized features. This is due to the statistical nature of the features. Rank 5 results for both normalized and non-normalized feature sets are shown in Table 5.2. The experimental results shown in Table 5.1 and 5.2 indicate that db4 performs slightly better than the db1 and db2 kernels and hence is more suitable for the analysis of area signals. Therefore, we choose db4 based non-normalized features for further experimental analysis.

We obtain 49.62% correct identification at rank 1 for probe 1 which increases to 66.69% at rank 5. The results for the probe 2 are better than the averaged results obtained for all the sequences. We achieve the CMS of 69.32% at rank 1 and 88.63% at rank 5 for the probe 2. This is due to the reason that both the gallery and probe sequences are from the same recording angle and viewpoint changes are minimal. For the probe 3, the rank 1 CMS drops to 26.67% and this value goes up to 71.11% at rank 5. The variation in the recording angle changes the viewpoint affecting the appearance of the person in the sequences. This degrades the performance of the system. In case of probe set 4, CMS value of 24.44% is obtained for the rank 1 while the rank 5 CMS is 60%. The variations in the appearance of the person in the silhouette sequence does
Table 5.1: Comparison of db1, db2 and db4 based features at rank 1

<table>
<thead>
<tr>
<th>Probe Set</th>
<th>Normalization</th>
<th>db1 (%)</th>
<th>db2 (%)</th>
<th>db4 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe 1</td>
<td>Raw</td>
<td>46.45</td>
<td>42.72</td>
<td>49.62</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>45.32</td>
<td>50.08</td>
<td>48.31</td>
</tr>
<tr>
<td>Probe 2</td>
<td>Raw</td>
<td>64.82</td>
<td>58.00</td>
<td>69.32</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>65.96</td>
<td>62.55</td>
<td>69.32</td>
</tr>
<tr>
<td>Probe 3</td>
<td>Raw</td>
<td>33.33</td>
<td>26.67</td>
<td>26.67</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>26.67</td>
<td>26.67</td>
<td>26.67</td>
</tr>
<tr>
<td>Probe 4</td>
<td>Raw</td>
<td>18.33</td>
<td>24.44</td>
<td>24.44</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>18.33</td>
<td>26.67</td>
<td>26.67</td>
</tr>
<tr>
<td>Probe 5</td>
<td>Raw</td>
<td>14.44</td>
<td>14.44</td>
<td>14.44</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>14.44</td>
<td>14.44</td>
<td>14.44</td>
</tr>
</tbody>
</table>

not change much due to the far angle. The lowest efficiency is obtained in case of probe set 5. We achieve a performance of 14.44% at rank 1 and 46.67% at rank 5 in terms of CMS. The reason for this drastic drop in CMS values is due to outdoor conditions. The appearance of shadows degrades the quality of the extracted silhouettes. Moreover, the lighting conditions and background also changes in the outdoor and contribute to more errors in segmentation thus degrading the quality of silhouettes further.

5.3.2 Choosing Level of Decomposition

Figure 5.3 shows the CMC curves for different levels of wavelet decomposition. The analyzing wavelet is db4 in all cases. The recognition performance increases as we increase the level of decomposition because the number of features increases. It is observed that beyond level 3 the increase in the recognition performance is not significant. Therefore we choose 3-level decomposition for further experimental evaluation.
### Table 5.2: Comparison of db1, db2 and db4 based features at rank 5

<table>
<thead>
<tr>
<th>Probe Set</th>
<th>Normalization</th>
<th>db1 (%)</th>
<th>db2 (%)</th>
<th>db4 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe 1</td>
<td>Raw</td>
<td>65.19</td>
<td>66.69</td>
<td>66.69</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>65.57</td>
<td>65.67</td>
<td>66.19</td>
</tr>
<tr>
<td>Probe 2</td>
<td>Raw</td>
<td>88.63</td>
<td>88.63</td>
<td>88.63</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>88.63</td>
<td>88.63</td>
<td>88.63</td>
</tr>
<tr>
<td>Probe 3</td>
<td>Raw</td>
<td>64.67</td>
<td>71.11</td>
<td>71.11</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>64.67</td>
<td>71.11</td>
<td>71.11</td>
</tr>
<tr>
<td>Probe 4</td>
<td>Raw</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Probe 5</td>
<td>Raw</td>
<td>46.67</td>
<td>46.67</td>
<td>46.67</td>
</tr>
<tr>
<td></td>
<td>Normalized</td>
<td>47.33</td>
<td>47.33</td>
<td>42.22</td>
</tr>
</tbody>
</table>

#### 5.3.3 Identification Results

We compare the performance of our method with the Baseline algorithm, GTech method and mass vector method. When comparing with the Baseline algorithm, we have not implemented the silhouette extraction part of the algorithm because we work directly with the silhouettes. Mass vector technique also works directly with the silhouettes. The similarity values for the Baseline algorithm are calculated from the silhouettes available in the GTech database. Table 5.3 summarizes the results of comparison for probe 1. The symbol $N$ appearing in the Table represents the number of frames in a silhouette sequence. It is observed from Table 5.3 that the proposed DWT method outperforms the Baseline method at rank 1 by 25.37%, however, its performance is lower than mass vector method by 23.89%. At rank 5, our method performs 16.32% better than the Baseline algorithm and significantly under performs mass vector method by 25.47%.

In Table 5.4, we provide the identification results for probe 2. The performance of our
method is 27.27% higher than the Baseline algorithm at rank 1 and significantly lower than mass vector method which performs perfectly and gives 100% recognition rate. Similarly mass vector method achieves perfect CMS value of 100% at rank 5 outperforming DWT and Baseline methods by a big margin. Baseline algorithm performs the lowest at rank 5 and its CMS value is 29.57% lower than our proposed DWT method.

In Table 5.5, we give the performance of the four techniques for probe 3. In this case, the performance of our DWT method is 4.45% higher than the Baseline algorithm at

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT method (db4)</td>
<td>16</td>
<td>49.62</td>
<td>66.69</td>
</tr>
<tr>
<td>Baseline silhouettes</td>
<td></td>
<td>24.25</td>
<td>50.37</td>
</tr>
<tr>
<td>Mass Vector [72]</td>
<td>128 x N</td>
<td>73.51</td>
<td>92.16</td>
</tr>
</tbody>
</table>

Figure 5.3: CMC curves for level 1 to level 5 decomposition
Table 5.4: Performance comparison for probe 2

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT method (db4)</td>
<td>16</td>
<td>69.32</td>
<td>88.63</td>
</tr>
<tr>
<td>Baseline silhouettes</td>
<td></td>
<td>42.05</td>
<td>59.09</td>
</tr>
<tr>
<td>Mass Vector [72]</td>
<td>128 x N</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.5: Performance comparison for probe 3

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT method (db4)</td>
<td>16</td>
<td>26.67</td>
<td>71.11</td>
</tr>
<tr>
<td>Baseline silhouettes</td>
<td></td>
<td>22.22</td>
<td>42.22</td>
</tr>
<tr>
<td>Mass Vector [72]</td>
<td>128 x N</td>
<td>97.78</td>
<td>100</td>
</tr>
<tr>
<td>GTech method</td>
<td>4</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

rank 1 and 28.89% better than the Baseline algorithm at rank 5. Mass vector and GTech method achieve much higher recognition rate than the propose technique both at rank 1 and rank 5. The CMS values for rank 2 are 100% for both mass vector and GTech methods.

Tables 5.6 and 5.7 summarize the results for the probe 4 and probe 5. The results are consistent with probe 3 with Baseline algorithm under performing the DWT method while GTech and mass vector methods achieving higher CMS values. The results for both GTech and mass vector are 100% at rank 5. The probe 5 results are the worst of all probes. This is due to the reason that the probe set consists of sequences which have been recorded outdoor while the gallery set has been recorded inside. Both GTech and mass vector method perform better than the proposed DWT based features both at rank 1 and rank 5. However, the performance of our method is similar to Baseline for
this probe. The reason for the better performance of GTech method over mass vector is the appearance of shadows in the silhouettes due to outdoor recording conditions. The GTech method is less sensitive to the shadows because it extracts body static parameters and stride parameters based on a coarse body modeling into regions.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>No. of Features (%)</th>
<th>Rank 1 (%)</th>
<th>Rank 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT method (db4)</td>
<td>16</td>
<td>24.44</td>
<td>60</td>
</tr>
<tr>
<td>Baseline</td>
<td>silhouettes</td>
<td>11.11</td>
<td>46.67</td>
</tr>
<tr>
<td>Mass Vector [72]</td>
<td>128 x N</td>
<td>93.33</td>
<td>100</td>
</tr>
<tr>
<td>GTech method</td>
<td>4</td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

In terms of the computational efficiency, the Baseline algorithm has the highest computational load among the 4 compared methods because it uses all the silhouettes directly as features. The mass vector method also carries higher computational cost than our proposed method. The time complexity of the similarity calculation process in the Baseline algorithm is $O(MN^2)$ compared to $O(MN)$ for our proposed feature extraction method. Where $M$ is the number of pixels in a frame and $N$ is the number of frames in the silhouette sequence. The time complexity of the mass vector feature extraction process is
$O(MN)$ and the matching process is $O(N^2)$.

The number of frames ($N$) for silhouette sequences in GTech database vary from a minimum of 45 to a maximum of 209 with a median value of 106. The typical size of the silhouette frames in GTech database is 240x400 pixels. Therefore the required number of floating point operations for the similarity calculation in case of Baseline algorithm will be of the order of 1 billion. In contrast the number of floating point operations required for our feature extraction method will be about 100 times smaller. We implemented the algorithms using 64 bit edition of Matlab 2011b ver. 7.13.0.564. We calculated the average execution time for feature extraction on a PC with Intel® Core i5-3210M @ 2.5GHz processor and 6GB RAM running 64 bit edition of Microsoft Windows 7.0 Home Premium Service Pack 1. The average time taken by the mass vector feature extraction method was 1.68 s for one silhouette sequence. The average execution time for the proposed DWT approach was 1.014 s. The Baseline algorithm does not extract any features and uses all the silhouettes as feature. Obviously there was no execution time involved in case of Baseline algorithm. However the recognition stage of Baseline algorithm is found to be very slow and the similarity calculation between one probe sequence and gallery sequence took an average of 52.027 s. Our proposed DWT approach performed matching between gallery and probe features in an average of 26.6 ms. The mass vector recognition stage took an average of 11.016 s per probe.

In terms of the compactness of the gait signature, our method provides a very compact signature which is just 16 dimensional. The most compact representation of gait is provided by GTech method which is just 4 dimensional although the method is a model based. The Baseline algorithm uses the whole silhouette sequence as a feature while mass vector method uses $N$ mass vectors each of length 128 as features. In the Baseline algorithm, the correlation of each frame from the partitioned probe sequences and gallery sequences is calculated for similarity measurements. This makes the Baseline algorithm very sensitive to the quality of silhouettes. This is the reason for low performance of the
algorithm when applied to the GTech silhouette database because of lower quality and noisy silhouettes.

5.4 Summary

This chapter presented a new method for human gait recognition. The proposed DWT based gait feature vector is low dimensional and comparatively easier to extract from the binary silhouette sequences. The proposed gait signature performs much better than the Baseline algorithm on the GTech silhouette database. It is also observed that the quality of the silhouettes is critical for the performance of the system. The proposed method is very sensitive to the quality of silhouettes and degrades drastically with the change in viewpoint and shadows in the silhouettes. Similarly the Baseline algorithm is very sensitive to the silhouette quality too and its performance is the lowest among the 4 compared methods.

The variation in the viewpoint affects the performance of the system. This variation is inherent to all appearance based systems because changing the viewpoint also changes the appearance of the human subjects in the image sequences. The mass vector method performs better than the proposed method for all probe sets. However when compared in terms of the compactness of the gait signatures and time complexity, the proposed method is superior to both Baseline and mass vector methods and compares well with GTech method. The mass vector method uses a very high dimensional gait signature and same is the case for Baseline algorithm.
Chapter 6

Conclusion and Future Work

6.1 Conclusion

Feature extraction is the most critical step in the human gait recognition systems. The dependence of gait on different joints and joint angles has been studied extensively in the medical and rehabilitation research community. Although gait is a dynamic process, yet our perception of gait comes both from dynamic as well as static information. The relationship of static and dynamic information and their relevant importance caught the attention of researchers in the past. But we found a lack of work in comparative performance analysis of features from different parts of the silhouettes in an appearance based setup. The leg and arm motion is the most pronounced during the walking process. We attempted to compare the dynamics of leg and arm for gait recognition. Our study confirmed the general conception of leg motion being the most important determining factor in gait recognition. It was also interesting to note that shoulder motion plays more significant role in the recognition process for certain databases. The usefulness of the study hinges on the fact that lower parts of the leg are generally more noisy due to a variety of variations such as walking surface, occlusion and shadows. More robust and cleaner features may be extracted in such situations from the upper part of the
silhouettes. In other situations they can play a complementary role in the recognition systems.

The variation in the viewpoint and the time lapse between the gallery and probe sequence are also major variates in gait recognition. The variations in gait pattern over time is an established fact and will be harder to overcome. The fusion of gait features with other biometrics such as face can resolve this issue. Another route that may be adopted in such situations will be the inclusion of human input at some level. The system can short list a small number of candidates and a human expert can then make the final decision. This will be especially useful in some application where the input database is big. Criminal investigations is one of such scenarios. The variation in viewpoint will need some type of normalization of the pose.

This dissertation also proposed two new feature extraction methods for gait recognition. The new methods use silhouette area signals which are very easy and simple to measure. The correlation analysis of these area signals revealed that a significant performance increase can be achieved by using the new features over the benchmark Baseline method. Comparable recognition results were achieved in regards with other techniques with these features. The 1-D diagonal slice of the 3rd. order autocorrelation of the area signal was also used in our experiments but it failed to provide better performance. The simplicity and compactness of these new correlation based features is their major advantage and entails low computational overhead.

The performance of wavelet based features presented in chapter 5 is not at the desirable level. Although it performed consistently higher than the Baseline benchmark, its performance was way below the state of the art techniques. Nevertheless the DWT features has the potential due to its discriminatory power for gait recognition. DWT of area signals is very sensitive to different variates in the database such as viewpoint, lighting conditions and appearance of shadows. Its performance degrades drastically with variations in the above mentioned factors and they are not very robust to be used as a
single feature set. They can be combined with other more discriminatory features to achieve improvements in overall recognition results for the gait recognition systems.

6.2 Future Research Extension

The fundamental principle for appearance based gait recognition is some type of shape analysis during the gait walking process. The extracted features are dependent on the changes in the appearance of the subjects as they walk on their trajectory. Therefore these methods have high dependency on the quality of the silhouettes. The quality of the silhouette data also plays an important role in the model based approaches. However, model based approaches are generally less sensitive to small changes in appearance compared to their appearance based counterpart. The performance and robustness of the gait features can be improved by building some kind of quality measure into the system. The low quality silhouettes are a result of various factors during the recording process and imperfections in the silhouette extraction methodology. We are considering to extend this work to include a quality measure in the system. The quality measure for silhouettes in regards to their capability for providing a robust gait feature will decide their inclusion or exclusion in the feature extraction process. Another useful way to approach this inherent problem in appearance based gait recognition will be to improve the quality of the silhouettes. Our observations from the experimental analysis and the databases leads us to believe that the quality of some silhouettes is so low that disregarding them during feature extraction may be the only possibility. In some cases, the silhouette extraction algorithms have to be modified to achieve this purpose. Although this dissertation did not discuss the silhouette extraction process, we are considering to implement one as a future extension to the present work.

The appearance based features of the same person show larger variations with viewpoint. The general recording framework for video gait data is based on a fixed camera and
subjects walking in front of it in different directions. This is also the case in surveillance cameras where the persons move in all different directions. The appearance of the person in the image varies hugely with these viewpoint changes due to occlusion and body limbs. Synthesizing a normalized canonical pose can neutralize these variations and enhance the performance of the system significantly. We are considering to improve the performance of our proposed feature extraction methods by implementing pose normalization.
Huang et. al. proposed EMD to decompose non linear non stationary data into oscillatory modes called Intrinsic Mode Functions (IMF) [66]. The method separates IMFs from signals modulated in both amplitude and frequency. IMF is a function that satisfies two conditions:

- The number of extrema and the number of zero crossings must in an IMF are either equal or at most differ by one.
- The mean value of the envelope traced by the local maxima and the envelope defined by the local minima is zero.

EMD is a data driven technique which does not assume any pre-defined basis functions and is thus very adaptive in nature. EMD is based on the following assumptions:

1. The signal has at least two extrema, one maximum and one minimum.
2. The characteristic time scale is defined by the time lapse between the extrema.
3. If the signal does not have any extrema and only contains inflection points, then
it can be differentiated once or more times to reveal the extrema. The results are then obtained by integration of the components.

EMD method does not require a mean or zero reference and can be applied directly to signals with non-zero mean. IMFs are extracted by the sifting process which is applied iteratively until a predefined condition is satisfied or the residue becomes a monotonic function. The signal $x(t)$ can be then represented in the following form.

$$x(t) = \sum_{i=1}^{k} e_i + r_k$$  \hspace{1cm} (A.1)

where $e_i$ denotes the $i$th extracted empirical mode and $r_k$ is the residue which is either a constant or mean trend. The sifting procedure to obtain the IMF is summarized in 11 steps as given in Algorithm 1.
Algorithm 2: EMD algorithm

1: Extract all local extrema.
2: Determine the upper envelope by connecting all the local maxima by cubic spline interpolation.
3: Determine the lower envelope by connecting all the local minima by cubic spline interpolation.
4: Calculate the mean envelope \( m_1 \) from upper and lower envelopes.
5: Calculate the first component \( h_1 \) as follows:

\[
h_1 = x(t) - m_1 \tag{A.2}
\]

6: Check if \( h_1 \) satisfies the IMF definition.
7: if yes:

\[
e_1 = h_1 \tag{A.3}
\]

8: Calculate residue as follows:

\[
r_1 = x(t) - h_1 \tag{A.4}
\]

9: Go to step 1 and repeat the sifting process to extract more IMFs treating \( r_1 \) as the input data.
10: if no: Calculate \( h_{11} \)

\[
h_{11} = h_1 - m_{11} \tag{A.5}
\]
11: Where \( m_{11} \) is the mean envelope of \( h_1 \). Repeat the sifting \( k \) times until the stoppage criterion is met to get \( h_{1k} \).

\[
h_{1k} = h_{1(k-1)} - m_{1k} \tag{A.6}
\]

\[
e_1 = h_{1k} \tag{A.7}
\]
Appendix B

Implementation Details

B.1 Baseline Algorithm

The binary silhouettes extracted by Baseline algorithm are available as part of the GC database. We are using these silhouette sequences in our experiments and hence only implement the recognition part of the algorithm. The summary of our implementation is given below.

**Step 1:** Gait period, $N_g$, is determined by counting the number of pixels in the bottom half of the silhouette in each frame of the sequence. The number of foreground pixels, $N_f$, is maximum when the two legs are farthest apart and drop to a minimum when the legs overlap. The median of the distances between minima is calculated by skipping every other minimum. This skipping strategy gives two estimates of the gait cycle. Gait period is estimated by averaging these two values.

**Step 2:** Similarity scores are computed by spatial-temporal correlation. Let us denote the probe sequence containing $M$ frames by $S_P = \{S_p(1), ..., S_P(M)\}$ and gallery sequence containing $N$ frames by $S_G = \{S_G(1), ..., S_G(N)\}$. Probe sequence is then partitioned into disjoint subsequences of $N_g$ contiguous frames. $N_g$ is the estimated gait period of the probe sequence calculated in step 1. The $kth$ partition of the Probe sequence is denoted
by $S_{Pk} = \{S_p(kN_g),...,S_p((k+1)N_g)\}$. Gallery sequence is not partitioned.

**Step 3:** Similarity between the gallery and probe sequence is then calculated by the following equations:

$$FrameSim(S_P(i), S_G(j)) = \frac{\text{Num}(S_P(i) \cap S_G(j))}{\text{Num}(S_P(i) \cup S_G(j))} \quad (B.1)$$

$$\text{Corr}(S_{Pk}, S_G(l)) = \sum_{j=0}^{N_g-1} FrameSim(S_P(k+j), S_G(l+j)) \quad (B.2)$$

$$\text{Sim}(S_P, S_G) = \text{Median}_k(\max_l \text{Corr}(S_{Pk}, S_G(l))) \quad (B.3)$$

### B.2 Mass Vector Method

All binary silhouettes are first normalized to the same height. After the height is normalized, the mass vectors and similarity measure is calculated by the following two steps:

**Step 1:** The mass vector along a given row is the number of pixels with a nonzero value in a that row of the silhouette and is calculated by the equation given below.

$$M(t) = [m_1(t), m_2(t), ..., m_H(t)] \in \mathbb{R}^H \quad (B.4)$$

$$m_h(t) = \sum_x I(x, y, t) \quad (B.5)$$

Where $I(x, y, t)$ is the pixel value (0 or 1) of binary silhouette at $t$th frame and $x$ and $y$ are the horizontal and vertical coordinates of the silhouette respectively.

**Step 2:** Matching is performed by DTW algorithm. The sequences are processed so that the first and the last frames are both rest stances. Euclidean distance is used as the local distance measure when comparing two mass vectors. The cumulative distance at the end of the warping path is recorded as the matching score between the reference and test samples. The classification is performed by the following equation:

$$\text{cls}(M) = \arg \min_i DTW(P, \tilde{M}_i) \quad (B.6)$$

where $\tilde{M}$ is the gallery sequence of person $i$ and $P$ is the probe sequence.
Appendix C

Wavelet Transform

C.1 Continuous Wavelet Transform

The continuous wavelet transform (CWT) was developed as an alternative to the STFT. The approach is similar to STFT in the sense that the function is multiplied with a wavelet similar to the window function in the STFT. However, there are several important differences between the two techniques:

1. The width of the window used by the Wavelet transform is not constant and changes as the transform is calculated for every single spectral component. This is the most significant characteristic of the wavelet transform.

2. Negative frequencies are not computed.

The continuous wavelet transform is defined as [75]:

\[ CWT(\tau, s) = \Psi(\tau, s) = \int_{-\infty}^{+\infty} x(t) \psi^*_\tau,s(t) dt \quad (C.1) \]

\[ \psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}} \psi\left(\frac{t - \tau}{s}\right) \quad (C.2) \]

where \( \tau \) and \( s \) are translation and scale parameters respectively and the superscript * refers to the complex conjugate. \( \psi(t) \) is known as the Mother wavelet that serves
Appendix C. Wavelet Transform

as a prototype for generating other window functions $\psi_{\tau,s}(t)$ known as the daughter wavelets. The daughter wavelets are obtained by shifting and scaling the mother wavelet. It must be noted here that the wavelets are finite length oscillatory functions. The translation process is similar to that of STFT; where the window function is moved over the entire signal. In the Wavelet transform, the scale parameter ($s$) replaces the frequency parameter. Large scales represent the global view of the signal or the low frequencies; while low scales represent the high frequencies. Mathematically speaking, large scales dilate the signal while low scales correspond to the compressed signals. CWT is a continuous transform and, therefore, $\tau$ and $s$ are incremented continuously. Since the transform is to be computed using a digital computer, both parameters are increased by a sufficiently small step size. This means that the time-scale plane is sampled and becomes discrete.

Continuous wavelet transform is a reversible transform subject to the constraint:

$$\{2\pi \int_{-\infty}^{\infty} \left| \hat{\psi}(\xi) \right|^2 d\xi \}^{\frac{1}{2}} < \infty$$

where $\hat{\psi}$ is the Fourier transform of $\psi$. The above equation implies that $\hat{\psi}(0) = 0$ or:

$$\int \psi(t) dt = 0$$

This is not a very restrictive condition and wavelet functions can be found that satisfy the above condition.

In the case of CWT analysis of the signal, the discretization process may be performed in any desired way. However, the Nyquist sampling rate is the minimum sampling rate required for the reconstruction/synthesis of the signal. Mathematically, the discretization process is defined by the following equation:

$$\psi_{n,k}(t) = s_0^{-n/2}\psi(s_0^{-n}t - k\tau_0)$$

where $s_0 > 1$ and $\tau_0 > 0$ are the discrete versions of scale and translation parameters while $n$ and $k$ are the step sizes for scale and translation respectively. The scale parameter
is discretized on a logarithmic grid while the time parameter is discretized based on the scale parameter. This means that the sampling rate for the time parameter is dependent on the value of the scale; and is different for different scales. The base of the logarithm depends on the user, 2 being the most common one. Only a finite number of points are taken. For example, if the base of logarithm is 2 then scale will have values 2, 4, 8, 16 and so on.

C.2 Discrete Wavelet Transform

A discrete version of the transform is required so that it can be computed using digital computers. CWT can be computed using a digital computer by discretizing the time-scale plane as shown in the previous section. However, discretized CWT still contains a lot of redundant information wasting a large amount of computational resource. DWT is defined to reduce the computational complexity. DWT provides sufficient information for signal analysis and reconstruction. DWT is derived by critically sampling CWT and is defined in the following equation.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{S_0^j}} \psi\left(\frac{t - k\tau_0}{S_0^j}\right)$$  \hspace{1cm} (C.6)

Where $S_0 = 2$ for sampling on the dyadic grid. The $\tau_0$ is usually taken as 1. $j$ and $k$ are integers and $t$ is the time variable. Therefore Equation C.6 can be written as follows.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k\tau_02^j}{2^j}\right)$$  \hspace{1cm} (C.7)

DWT is obtained by passing the signal through a series of low pass and high pass filters. When a signal is passed through a filter, the signal is convolved with the impulse response of the filter to produce the output signal. The filtering operation for a discrete signal is defined as follows:

$$y_{\text{low}}[n] = \sum_{k=\text{\text{-\infty}}}^{\infty} x(k)g(2n - k)$$  \hspace{1cm} (C.8)
where $y_{\text{high}}[n]$ is the output of the system with impulse response $g(n)$ and $x(n)$ is the input signal. $y_{\text{high}}[n]$ is the output of the system with impulse response $h(n)$. The filters are half band digital low pass filters. The output of these filters therefore contains only the frequencies up to half of the maximum frequency of the original signal. The frequencies higher than half of the maximum frequency in the original signal are removed by the low pass half band filter. Therefore, we can eliminate half of the samples by subsampling without any loss of information. This subsampling operation doubles the scale of the signal since half of the samples are now removed. The filtering operation, on the other hand, reduces the frequency resolution by removing half of the spectral components from the signal.

$h(n)$ is a low-pass filter and its corresponding mirror filter $g(n)$ is defined as:

$$g(n) = (-1)^n h(N - 1 - n)$$  \hspace{1cm} (C.10)

These filters $g(n)$ and $h(n)$ are called quadrature mirror filters (QMF). The discrete wavelet transform is implemented by a quadrature mirror filter bank.
Bibliography


