UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
School of Electronics and Computer Science

GAIT ANALYSIS AND RECOGNITION FOR AUTOMATED VISUAL SURVEILLANCE

by

Imed Bouchrika

A thesis submitted for the degree of
Doctor of Philosophy

June 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

GAIT ANALYSIS AND RECOGNITION FOR AUTOMATED VISUAL SURVEILLANCE

by Imed Bouchrika

Human motion analysis has received a great attention from researchers in the last decade due to its potential use in different applications such as automated visual surveillance. This field of research focuses on the perception and recognition of human activities, including people identification. We explore a new approach for walking pedestrian detection in an unconstrained outdoor environment. The proposed algorithm is based on gait motion as the rhythm of the footprint pattern of walking people is considered the stable and characteristic feature for the classification of moving objects. The novelty of our approach is motivated by the latest research for people identification using gait. The experimental results confirmed the robustness of our method to discriminate between single walking subject, groups of people and vehicles with a successful detection rate of 100%. Furthermore, the results revealed the potential of our method to extend visual surveillance systems to recognize walking people.

Furthermore, we propose a new approach to extract human joints (vertex positions) using a model-based method. The spatial templates describing the human gait motion are produced via gait analysis performed on data collected from manual labelling. The Elliptic Fourier Descriptors are used to represent the motion models in a parametric form. The heel strike data is exploited to reduce the dimensionality of the parametric models. People walk normal to the viewing plane, as major gait information is available in a sagittal view. The ankle, knee and hip joints are successfully extracted with high

accuracy for indoor and outdoor data. In this way, we have established a baseline analysis which can be deployed in recognition, marker-less analysis and other areas.

The experimental results confirmed the robustness of the model-based approach to recognise walking subjects with a correct classification rate of 95% using purely the dynamic features derived from the joint motion. Therefore, this confirms the early psychological theories claiming that the discriminative features for motion perception and people recognition are embedded in gait kinematics. Furthermore, to quantify the intrusive nature of gait recognition we explore the effects of the different covariate factors on the performance of gait recognition. The covariate factors include footwear, clothing, carrying conditions and walking speed. As far as the author can determine, this is the first major study of its kind in this field to analyse the covariate factors using a model-based method.

Contents

Abstract i
Table of Contents iii
List of Tables vii
List of Figures viii
Acknowledgements xii

1 Context and Contribution **1**
1.1 Context 1
1.2 Main Contributions 2
1.3 Thesis Overview 3
1.4 List of Publications 5

2 Introduction **7**
2.1 Gait for Visual Surveillance 8
2.2 Challenges 10
2.3 Related Work 11
2.3.1 Automated Visual Surveillance 11
2.3.2 Gait Analysis and Recognition 12
2.3.2.1 Non-Model Based Approach 13
2.3.2.2 Model-Based Approach 13
2.4 Conclusions 14

3 Human Motion Analysis **16**

3.1 Automated Motion Analysis System . . . 17
3.1.1 Marker-based Motion Analysis Systems . . . 17
3.1.2 Marker-less Motion Analysis Systems . . . 19
3.1.2.1 Tracking of Moving Objects . . . 20
3.1.2.2 High-Level Feature Extraction . . . 21
3.1.2.3 Classification Phase . . . 23
3.2 Human Motion Perception . . . 24
3.3 Gait for Human Motion Analysis . . . 26
3.3.1 Gait Analysis . . . 26
3.3.2 Human Gait Characteristics . . . 26
3.3.3 Human Gait Cycle . . . 28
3.3.4 Gait Angular kinematics . . . 30
3.3.5 Analysis of Joint Spatial motion . . . 30
3.4 Gait Databases . . . 32
3.4.1 Southampton (Inter-Subjects) Large database . . . 33
3.4.2 Southampton Covariate Database . . . 35
3.5 Conclusions . . . 35

4 Gait for People Detection **37**
4.1 Foreground Segmentation . . . 38
4.1.1 Background Subtraction . . . 38
4.1.2 Shadow and Noise Suppression . . . 39
4.2 Tracking Moving Objects . . . 40
4.2.1 Region Correspondence Algorithm . . . 42
4.2.2 Tracking Constraints . . . 43
4.2.3 Occlusion Handling . . . 45
4.3 Object Classification . . . 46
4.3.1 Heel Strike Extraction . . . 47
4.3.1.1 Computation of Proximity Image . . . 48
4.3.1.2 Peak Detection . . . 50

4.3.2 Derivation of Feature Vector . . . 51
4.4 Results and Analysis . . . 52
4.4.1 Tracking and Occlusion Analysis . . . 53
4.4.2 Classification Results . . . 54
4.4.3 Performance Analysis for Heel Strike Detection . . . 55
4.5 Conclusions . . . 57

5 Feature Extraction via Model-Based Method 59
5.1 Gait Periodicity Estimation . . . 60
5.2 Shape Extraction via Evidence Gathering . . . 62
5.2.1 The Hough Transform . . . 62
5.2.2 Shape Parametrization using Fourier Descriptors . . . 63
5.2.3 Recursive Evidence Gathering Algorithm . . . 65
5.3 Extracting the Joints' Positions . . . 67
5.3.1 The Ankle Joint . . . 68
5.3.2 The Knee and Hip Joint . . . 71
5.4 Experimental Results and Analysis . . . 73
5.4.1 Periodicity Detection . . . 73
5.4.2 Joint Extraction . . . 74
5.4.3 Performance Analysis . . . 78
5.5 Conclusions . . . 79

6 Gait Recognition by Dynamic Cues 81
6.1 Static vs. Dynamic Gait Features . . . 82
6.2 Gait Identification System . . . 84
6.3 Derivation of Gait Signature . . . 84
6.3.1 Extraction of Dynamic Features . . . 85
6.3.2 Fourier Description of Gait Features . . . 87
6.3.3 Normalisation of Gait Features . . . 88
6.4 Feature Selection and Classification Metrics . . . 89

6.4.1 Validation-Based Feature Selection . . . 90
6.4.2 Statistical-Based Feature Selection . . . 91
6.5 Gait Recognition and Performance Analysis . . . 92
6.5.1 Classification Results . . . 92
6.5.2 Statistical Analysis . . . 95
6.5.3 Gait Feature Analysis . . . 98
6.6 Conclusions . . . 100

7 Exploratory Factor Analysis of Gait Recognition 102
7.1 Covariate Factors in Gait . . . 103
7.2 Data Acquisition for Covariate Analysis . . . 104
7.2.1 Data Acquisition . . . 104
7.2.2 Gait Feature Extraction . . . 104
7.3 Covariate Analysis for Gait Recognition . . . 105
7.3.1 The Footwear Effects . . . 106
7.3.2 The Clothing Effects . . . 108
7.3.3 Load Carriage . . . 110
7.3.4 The Speed Effects . . . 112
7.4 Covariate Factor Analysis of Gait Features . . . 113
7.5 Conclusions . . . 114

8 Conclusions and Future Work 116
8.1 Conclusions . . . 116
8.2 Future Work . . . 118

A Parameters Reduction 120

B Anthrometric Measurements of the Human Body 123

References 124

List of Tables

3.1 Human Gait Databases. 33

4.1 Moving Objects Classification Results 55

6.1 Gait Recogntion using Static and Fusion of both Static and Dynamic Features. 83

6.2 Gait Recognition Results using Dynamic Cues 93

6.3 Mean and Standard deviation of Intra and Inter class distributions. 96

7.1 Statistical Analysis of Gait Features. 113

List of Figures

3.1 Marker-Based Systems for Human Motion Analysis. 18
3.2 Structure of a Marker-less Motion Capture System. 20
3.3 Model-Based Methods for Feature Extraction. 22
3.4 Frames taken from a Moving Light Displays video. 25
3.5 Human Gait cycle [131]. 28
3.6 Gait Angular Motion. 29
3.7 Spatial Motion Analysis of the Joints. 31
3.8 The Southampton Large Gait Database. 34
3.9 The Southampton Covariate Gait Database. 36

4.1 Foreground Segmentation 40
4.2 The Block Diagram for the Tracking Algorithm 44
4.3 Tracking Multiple Objects 46
4.4 The Heel Strike Regions during the Locomotion Process 47
4.5 Example Results for the Corner Proximity Measure 49
4.6 The corners proximity images for Single Walking Person, Group of People and Moving Vehicle 50
4.7 Extraction of Heel Strike Positions 51
4.8 Results of Tracking Multiple Moving Objects 53
4.9 Experimental Results for Handling Occlusion 54
4.10 Feature Space for Moving Object Classification 55
4.11 Experimental Results for Heel Strikes Extraction 56
4.12 Performance Analysis for Heel Strikes Detection. 57

4.13 Heel Strike Detection at Various Resolutions. 58

5.1 Gait Periodicity Detection : (a) The Start of the Terminal Stance Phase, (b) The smoothing of the Histogram function $P(t)$. 61
5.2 Parametric Representation of shapes using Elliptic Fourier Descriptors (EFD) 65
5.3 Extraction of Corner Points using the Skeletonization Process 69
5.4 Template model and horizontal displacement of the ankle motion 70
5.5 Global Pattern Extraction of the Ankle Joint. 70
5.6 Spatial Motion Templates: (a) Knee Joint. (b) Hip Joint. 71
5.7 Global Pattern Extraction of the Knee and Hip Joint. 72
5.8 Experimental Results for Gait Periodicity Detection 74
5.9 Derivation of Gait Angular Motion during one Gait Cycle 75
5.10 Automated Extraction of the Joints' Positions for Indoor Data 76
5.11 Automated Extraction of the Joints for Outdoor Data 77
5.12 Performance Analysis for The Joint Extraction. 78
5.13 The Extraction of Joints for Indoor Data with self-occlusion 80

6.1 Structure of Gait Identification System 85
6.2 Gait Angular Measurements 86
6.3 The Cumulative Match Score and ROC Curve for the Classification Results. 94
6.4 Correlation Matrices for Gait Recognition 96
6.5 Intra-Class and Inter-Class Distributions of Gait Signatures 96
6.6 Projection of Database I using Generalized Discriminant Analysis 97
6.7 Dynamic Feature Analysis for Gait Recognition 99
6.8 Analysis of Fourier Components for Gait Recognitions 100
6.9 Fusion Analysis for Gait Recognition. 101

7.1 Performance Analysis of Gait Recognition for Covariate Data. 106
7.2 Footwear Covariate Factors. 107
7.3 Classification Results for the Footwear Covariates. 108

7.4 Clothing Covariate Factors. 109
7.5 Classification Results for the Clothing Covariates. 110
7.6 Clothing Covariate Factors. 111
7.7 Classification Results for the Load Carriage. 112
7.8 Classification Results for the Walking Speed Covariates. 113
7.9 Angular Feature Analysis for Covariate Gait Recognition. 114
7.10 Displacement Feature Analysis for Covariate Gait Recognition. 115

B.1 Body segment properties [132]. 123

Declaration of Authorship

I, Imed Bouchrika, declare that the thesis entitled Gait Analysis and Recognition for Automated Visual Surveillance and the work presented in it are my own, I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as listed in Section 1.4 of this thesis.

Signed: _____

Date : _____

Acknowledgements

I would like to express my immense and special gratitude to my supervisor Prof. Mark Nixon for his help and guidance throughout the course of this research. I truly admire his unique way of teaching and supervising as well as his friendliness. Again million thanks are due to Prof. Nixon who introduced me to the field of Computer Vision. I also wish to acknowledge all people who helped me through this research either directly or indirectly including: Dr. John Carter for his valuable feedback during the mini-thesis exam, Xin Lui and all people from the ISIS research group. Thanks also go to people who had spent countless hours developing open-source software or setting up databases for the academic community.

I am especially grateful to the Algerian government for sponsoring my studies in the United Kingdom. I thank my parents (Rabah and Mehenia) for rising me up and praying for me to finish my studies successfully and return home safely. Also, thanks are due to my wife Zohra for her support during my PhD. Thanks go also to my baby daughter Asma for waking me up early every day!

To my parents....

Chapter 1

Context and Contribution

1.1 Context

Surveillance technology is now ubiquitous in modern society. This is due to the increasing number of crimes as well as the vital need to provide a safer environment. Because of the rapid growth of security cameras and impossibility of manpower to supervise them, the deployment of biometric technologies becomes important for the development of automated visual surveillance systems. Recently, the use of gait for people identification in surveillance applications has attracted researchers from computer vision. The suitability of gait recognition for surveillance systems emerges from the fact that gait can be perceived from a distance as well as its non-invasive nature. Interestingly, in one of the high profile murder cases in the UK where a child was abducted and killed, the identity of the murderer could not be revealed directly from the surveillance video footage. The only solution that could be employed to determine the suspect's identity in this situation was gait recognition, as proposed by researchers from the University of Southampton [92]. However, gait as a biometric is still in its infancy and most of the gait recognition methods rely on whole body analysis.

1.2 Main Contributions

A number of new methods and approaches based on gait analysis and suited for deployment in automated visual surveillance applications are proposed in this study. The major contributions of this research are presented in this section. We propose a layering approach to ease the process of tracking multiple moving objects based on low-level feature correspondence between consecutive frames [D]. Moving objects are assigned to different layers whereby blobs corresponding to the same object are assigned to the same layer. The criteria for allocating objects to layers are based on the Mahalanobis distance measure of shape-based features. A number of constraints is imposed to efficiently handle occlusion, background clutter and other difficulties encountered during the tracking process.

Because of the dearth of visual surveillance systems that exploit human gait for object classification and their limited aim of only detecting pedestrians using simple shape-based or motion-based features extracted from silhouette data, we have explored an alternative technique for walking people detection and recognition based on their gait pattern [A,B]. In our method, walking pedestrians are distinguished from other moving objects via exploiting the rhythmic and periodic nature of human gait. The foot print pattern produced from the heel strikes of walking subjects is considered the primary cue for people detection.

Furthermore, we present a new method aimed to extract the foot print pattern of walking subjects (i.e. the heel strike locations) [E]. The method is based on the observation that the foot is stabilised on the ground for around half a gait cycle during the striking phase. Therefore, if we extract corners or edges of every frame and combine them together, dense or crowded areas will be produced at the heel strike positions. In order to derive the heel strikes, a new measure of point proximity is introduced in this research which gives a discrete value for the density of points in an image. The proximity value for a given point is evaluated based on the number of points in the neighbourhood region and their distance to the point.

A new model-based method is described to extract the positions of the joints for walking people [C]. A recursive evidence gathering algorithm is presented for the extraction process whereby spatial model templates for the human motion are described in a parameterized form using the Elliptic Fourier Descriptors. In this way, we have established a baseline analysis which can be deployed in recognition, marker-less analysis and other areas. Moreover, a novel approach for developing an automated non-invasive biometric system based on gait is described. The gait signature is derived using the model-based method. As most surveillance systems are limited to tracking and detecting moving objects, we have made a new attempt to extend visual surveillance systems to recognise walking people using a biometric solution based on gait.

Finally, intensive research is carried out to confirm the early psychological theories claiming that the discriminative features for motion perception and people recognition are embedded in gait kinematics. We show that the gait angular measurements derived from the joint motions possess most of the discriminatory power for gait recognition. Furthermore, we explore the effects of the different covariate factors of gait recognition including footwear, clothing, carrying conditions and walking speed. This is the first major study of its kind in this field which attempts to analyse the covariate factors using a model-based method.

1.3 Thesis Overview

This thesis is structured as follows:

- Chapter 2: ***Introduction***

  In this chapter, an introductory motivation for the use of gait as a biometric method and its prospect deployment in visual surveillance applications is presented. This is followed by a discussion of the various challenges encountered to develop a gait recognition system for smart surveillance. Finally, the chapter extensively reviews the research carried out in the field of visual surveillance with a particular emphasis on gait recognition.

- Chapter 3: ***Human Motion Analysis***

  This chapter is devoted to providing an insight into human motion perception and gait analysis. The different motion capture systems employed for motion analysis including marker-based and marker-less are reviewed. This is followed by a discussion of the different stages involved in the processing and perception of human motion using vision-based systems. The chapter delves into the arena of gait analysis describing the gait cycle, gait characteristics and gait motion in full details. Lastly, a brief review of the current gait databases is plotted.

- Chapter 4: ***Gait for People Detection***

  In this chapter, the proposed method for tracking multiple moving objects in an unconstrained environment is described. The method uses a layering technique in order to establish motion correspondence between frames based on low-level features. The chapter presents the gait-based approach employed for the classification of moving objects. The rhythmic pattern of gait motion derived via the extraction of the heel strikes is utilised as the main cue to distinguish walking subjects from other moving objects. The experimental results for tracking and classifying moving objects are discussed at the end of the chapter.

- Chapter 5: ***Feature Extraction via Model-Based Method***

  This chapter is dedicated to presenting the proposed model-based method for the extraction of the joint trajectories of walking subjects. The spatial templates describing the human gait motion are produced via gait analysis performed on data collected from manual labelling. The Elliptic Fourier Descriptors are used to represent the motion models in a parametric form. We show how the heel strike data can be exploited to reduce the dimensionality of the parametric models. In this chapter, the recursive evidence gathering algorithm used for the extraction of the joints is described. The experimental results for the extraction of the joints in indoor and outdoor scenarios are presented at the end of the chapter.

- Chapter 6: ***Gait Recognition by Dynamic Cues***

  In this chapter, the different types of features used for gait analysis and recognition

are presented citing some of the reported results for gait recognition using body-related features. The different stages for the derivation of gait signatures for walking subjects using dynamic measurements are described. The recognition significance of the various dynamic cues which are extracted from the joint angular motion of the lower limbs is examined for constructing a gait signature based purely on the gait kinematics.

- Chapter 7: ***Exploratory Analysis of Gait Recognition***
  This chapter is devoted to investigating the covariate effects on gait recognition. We present a brief review from the medical literature describing the various factors which affect the gait pattern. Afterwards, experimental studies are carried out to explore the different covariates on gait recognition using dynamic features. The covariate factors include footwear, clothing, carrying conditions and walking speed.

- Chapter 8: ***Conclusions and Future Work***
  Overall conclusions are drawn in this chapter. Future work and research thrusts are discussed at the end of the chapter.

1.4 List of Publications

So far the following papers have been published:

[A] I. Bouchrika and M. S. Nixon. People Detection using Gait for Visual Surveillance, *BMVA Symposium on Detection vs. Tracking*, London, United Kingdom 2006

[B] I. Bouchrika and M. S. Nixon. People Detection and Recognition using Gait for Automated Visual Surveillance, in *Proceedings of the IEE International Symposium on Imaging for Crime Detection and Prevention*, pp.576-581, London, United Kingdom, June 2006

[C] I. Bouchrika and M. S. Nixon. Markerless Feature Extraction for Gait Analysis, in *Proceedings of 5th IEEE SMC Chapter Conference on Advances in Cybernetic Systems*, pp.55-60, Sheffield, United Kingdom. September 2006

[D] I. Bouchrika and M. S. Nixon. Gait-Based Pedestrian Detection for Automated Surveillance. in *Proceedings of International Conference on Computer Vision Systems*, Bielfeld, Germany, March 2007

[E] I. Bouchrika and M. S. Nixon. Model-Based Feature Extraction for Gait Analysis and Recognition, in *Proceedings of Mirage: Computer Vision / Computer Graphics Collaboration Techniques and Applications* pp.150-160, INRIA Rocquencourt, France, March 2007

[F] I. Bouchrika and M. S. Nixon. Exploratory Factor Analysis of Gait Recognition, in *Proceedings of 8th International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, September 2008

Chapter 2

Introduction

In recent years, automatic visual surveillance has received considerable interest in the computer vision community. This is due to the increasing numbers of crimes from robbery to terrorist attacks, as well as the inability of human operators to monitor the increasingly growing numbers of surveillance cameras deployed in security sensitive areas such as government buildings and airports, or public places such as shopping malls and streets. According to the British Security Industry Association, the number of CCTV cameras installed in the UK was estimated to be more than 4.25 million in 2004; this figure is expected to grow rapidly particularly after the terrorist attacks that London witnessed in July 2005. Despite the huge increase of surveillance systems, the question whether current surveillance systems work as a deterrent to crime is still debatable. Security systems should not only be able to predict when a crime is about to happen but, more importantly, they ought to identify the individuals suspected of committing crimes, say through the use of biometrics such as gait recognition. In this chapter, the use of gait analysis for visual surveillance is introduced. This is followed by a discussion of the various challenges in developing a smart surveillance system. Finally, the chapter extensively reviews the research carried out in the field of visual surveillance with a particular emphasis on gait recognition.

2.1 Gait for Visual Surveillance

A vision-based system for automated surveillance consists of three main phases: detection, tracking and perception. In the last phase, a high-level description is produced based on the features extracted during the previous phases from the temporal video stream. The main aim of a smart surveillance system is to detect and track people in the scene as well as to perceive their behaviour and report any suspicious activities to the control centre. However, an ideal security system for crime detection and prevention should be able to recognise potential offenders from existing data.

Identification systems will undoubtedly play a key role in aiding law enforcement officers in their forensic investigations. More importantly, by early recognition of suspicious individuals who may pose security threats, the system would be able to reduce future crimes. Human motion perception has been of interest to researchers from different disciplines due to the wide range of applications ranging from activity recognition to people identification. In fact, early studies by Johansson [63] on human motion perception using Moving Light Displays (MLD) have revealed that an observer can recognise different types of human motion based on joint motions. Moreover, the observer can make a judgement of the gender of the person [70], and even further identify the person if they are already familiar with their gait [45]. This leads to the conclusion that gait might be a potential biometric for surveillance systems.

A *biometric* is a descriptive measure based on the human behavioural or physiological characteristics [59] which distinguishes a person uniquely among other people; this unique description should be universal and permanent. Currently, as most biometric systems are still in their infancy [59], the use of biometrics is limited to identity verification and authentication. Gait is an emergent biometric which is increasingly attracting the interests of researchers as well as the industry. Gait is defined as the manner of locomotion, i.e. the way of walking. Early studies by Murray [84] revealed that gait might be a useful biometric for people identification, a total of 20 feature components including ankle rotation, spatial displacement and vertical tipping of the trunk have been identified to render uniquely the gait signature for every individual. While some of

these features are difficult to extract using current computer vision systems, others are not consistent over time for the same person [84]. In one of the early experiments on gait recognition conducted by Cutting *et al* in 1978 [70], it was demonstrated that people can recognise others just by gait cues. Interestingly, in one of Shakespeare's theatre plays (The Tempest: Act 4, Scene 1) the following sentence gives a clear indication that gait recognition is not a novel concept:

"High'st Queen of state, Great Juno comes;
I know her by her gait".

Although gait recognition is still a new biometric and is not sufficiently mature to be deployed in real world applications such as visual surveillance, it has the potential to overcome most of the limitations that other biometrics suffer from such as face, fingerprints and iris recognition which can be obscured in most situations where serious crimes are involved. Face recognition in many cases has been proven to be unreliable for visual surveillance systems; this is due to the fact that people can disguise or hide their faces as well as that video data being captured can be inadequate at low resolution. Furthermore, another major drawback of face identification in security applications is its low recognition rates in poor illumination. This is because most of the facial features cannot be recovered at large distances even using night vision capability [66]. Although fingerprint and iris recognition have proved to be robust for applications where authentication or verification are required, such biometrics are inapplicable for situations where the subject's consent and cooperation are impossible to obtain.

Unlike other biometrics which might not be perceivable at low resolution, gait recognition has attracted interest as it has the potential to be efficient even if the individual being recognised is at a distance from the camera. Another advantage of gait recognition is that it is a non-invasive biometric and hence does not require the subject to cooperate with the system. Furthermore, as the purpose of any reliable biometric system is to be robust enough to reduce the possibility of signature forgery, a gait signature which is based on human motion is hard to conceal and forge. Recently, Lynnerup *et al* [79] affirmed the usefulness of gait analysis in forensics. They were able to identify two bank

robbers by matching surveillance images with images of the suspects and this evidence was later used to convict two suspects. Based on body features, gait and anthropometric[1] measures, Lynnerup argued that there was a strong resemblance between the suspect and one of the perpetrators.

2.2 Challenges

Despite the recent outstanding advancements in computer vision and pattern recognition technologies, there are still major challenges to be overcome for the realisation of automated visual surveillance. Accurate and robust segmentation as well as the tracking of multiple moving objects in an unconstrained, dynamic and cluttered environment are only a few of the numerous difficult challenges. Moreover, people detection and moving object classification using a single surveillance camera is an intricate task. The difficulties stem from a number of factors related either to the environment such as illumination changes, shadows and occlusion [74] or to the nature of human being such as self-occlusion, articulation and appearance variation due to the clothing type [1].

Most recent vision systems for human motion analysis rely primarily on the extraction of markers attached to the joints of moving people. However, most applications such as visual surveillance require the deployment of an automated marker-less vision system to recover the joints' trajectories. On the other hand, automated extraction of the joints' positions from videos of walking subjects is still an unsolved problem [126, 76] as non-rigid human motion encompasses a wide range of possible motion transformations due to the highly flexible structure of the human body and to self occlusion [139, 43]. Furthermore, clothing type, segmentation errors and different viewpoints pose a substantial challenge for accurate joint localisation.

[1] *Anthropometrics*: The systematic quantitative representation of the human body. Anthropometric techniques are used to measure the absolute and relative variability in size and shape of the human body. McGraw-Hill Encyclopedia of Science and Technology, 2005

It becomes clear that developing a biometric system based on gait motion is a highly demanding problem in computer vision. This is due to the previously mentioned difficulties encountered during the extraction of gait motion and other additional and equally critical factors to consider what might affect the derivation of gait signature, such as the physical and physiological conditions of the subject [108, 91]. Finally, the realization and deployment of automated marker-less gait analysis and recognition in real world applications would be a major contribution for various disciplines including forensics, security and medical applications.

2.3 Related Work

Much research in computer vision is directed to the analysis of articulated moving objects and more specifically into human gait analysis. This research is fuelled by the wide range of applications where gait analysis can be deployed, such as [55] smart visual surveillance and athletic performance. There is a large collection of literature from various disciplines that approves of the concept of using gait for people recognition. The following section sheds light on the recent state of the art studies related to gait analysis and recognition covering the different methodologies employed for feature extraction. As a potential application for gait analysis, we review the latest research for automated visual surveillance based on human motion analysis.

2.3.1 Automated Visual Surveillance

Existing surveillance systems have been classified into several categories [51] according to their type (single or multiple camera) and their functionality (tracking single , multiple people, etc.). Wren *et al* [136] proposed the PFinder system. It uses a unimodal background model to locate moving objects. The proposed system however is constrained to analysing scenes containing a single individual at a time, which may not be realistic for real world circumstances. The W^4 [51] surveillance system employs an appearance model to track people whereby single subject or groups are distinguished

using a projection histogram. Each person in the group is tracked via the detection of their head. Lipton *et al* [77] proposed a real time vision-based system to classify moving objects into either human or vehicle based on the "dispersedness" i.e. the compactness of the shaped computed as the ratio of the perimeter squared to the area. In their work, people are assumed to have a dispersedness value smaller than vehicles, however shape metrics can vary depending on image size and distance from camera.

Wang [55] surveyed two types of different features used for people detection in surveillance systems: shape-based and motion-based cues. The first type relies on the shape of human silhouettes such as dispersedness [77], aspect ratio of the bounding box, or just simple shape parameters. For the motion-based features, the periodicity of human motion is considered as a strong cue for people detection. Cutler [28] described a real time method for measuring motion periodicity based on self-similarity, which is used to distinguish walking people from other moving objects. On the other hand, Javed *et al* [62] proposed a different motion-based feature which is based on the rigidity and self-articulation nature of moving objects. The motion measurement named Recurrent Motion Image (RMI) computes the repeated internal motion to classify moving objects into single person, vehicle or group of subjects.

2.3.2 Gait Analysis and Recognition

Much of the interest in the field of human gait analysis has originated from physical therapy, orthopaedics and rehabilitation practitioners for the diagnosis and treatment of patients with walking abnormalities. As gait has recently emerged as an attractive biometric, gait analysis has become a challenging computer vision problem. Many research studies have aimed to develop a system capable of overcoming the difficulties imposed by the extraction and tracking of human motion features. Various methods were surveyed in [43], [81], [124] and [3]. Aggarwal *et al.* [2] categorised the different vision-based methods used for human motion analysis into non-model based and model-based methods. In what follows, both approaches with their related work are outlined.

2.3.2.1 Non-Model Based Approach

In the non-model based approach, feature correspondence between successive frames is based upon prediction, velocity, shape, texture and colour. Shio *et al.* [106] proposed a method to describe the human body using moving blobs or 2D ribbons. The blobs are grouped based on the magnitude and the direction of the pixel velocity. Kurakake and Nevatia [73] worked on the extraction of joint locations by establishing correspondence between extracted blobs. Small motion between consecutive frames is the main assumption, whereby feature correspondence is conducted using various geometric constraints. For the deployment of non-model based methods for gait recognition, Little *et al.* [78] derived a gait signature using dense optical flow where they achieved a recognition rate of 95% on a database of 6 subjects. Huang [56] proposed a new approach for automatic gait recognition based on statistical analysis. The gait signature for every subject was produced by combining the eigenspace transformation (EST) with the canonical space transformation (CST) in order to reduce the dimensionality of the silhouette data and optimise the class separability respectively.

2.3.2.2 Model-Based Approach

For the model-based approach, a prior shape model is established to match real images to this predefined model, and thereby extracting the corresponding features once the best match is obtained. The stick and volumetric models [139] are the most commonly used methods. Akita [7] proposed a model consisting of six segments comprising of two arms, two legs, the torso and the head. Guo *et al* [49] represented the human body structure by a stick figure model which had ten articulated sticks connected with six joints. Rohr [104] proposed a volumetric model for the analysis of human motion using 14 elliptical cylinders to model the human body. Recently, Karaulova *et al.* [67] used the stick figure model to build a novel hierarchical model of human dynamics represented using Hidden Markov Models. The model-based approach is the most popular method for human motion analysis due to its various advantages [87]: it can extract detailed and accurate motion data, and is capable of coping well with occlusion and self-occlusion.

A few number of model-based methods has been proposed for gait recognition. Niyogi *et al* [93] were perhaps the pioneers in deploying a model-based method for gait recognition. Gait signature is derived from the spatio-temporal pattern of a walking subject using a five stick model. Using a database of 26 sequences containing 5 different subjects, a promising classification rate of 80% was achieved. Cunado [27] used the Velocity Hough Transform to extract the hip angular motion via modelling human gait as moving pendulum. The gait signature is derived as the phase-weighted magnitudes of the Fourier components. A recognition rate of 90% was achieved using the derived signature on a database containing 10 subjects. Yam *et al.* [138] modelled the human gait as a dynamic coupled oscillator which was used to extract the hip and knee angular motion via evidence gathering. The method was evaluated on a database of 20 walking and running subjects, achieving a recognition rate of 91% based on gait signature derived from the Fourier analysis of the angular motion. In a more recent work, Wagg *et al.* [122] proposed a new model-based method for gait recognition based on the biomechanical analysis of walking subjects. Mean model templates are adopted to fit individual subjects. Both the anatomical knowledge of human body and hierarchy of shapes are used to reduce the computational costs. The gait feature vector is weighted using statistical analysis methods to measure the discriminatory potency of each feature. On the evaluation of this method, a correct classification rate of 95% is achieved on a large database of 2163 video sequences of 115 different subjects.

2.4 Conclusions

Gait is an emergent biometric and recent studies confirmed its potential use for surveillance applications. As computer vision researchers approached the problem of gait analysis and recognition using different methodologies including model-based or model-free methods, most of their contributions and research studies were limited to the use of a silhouette-based approach or anatomical-based methods for gait recognition applied to walking subjects recorded from the side view without examining the effects of everyday factors including clothing, load carriage and footwear. We propose a model-based

method for the recovery of joint trajectories. This is because the model-based approach is more suited to general deployment. Furthermore, we have investigated the use of gait motion for biometric applications and the effects of various factors on the performance of gait recognition.

Chapter 3

Human Motion Analysis

The perception of human motion is one of the most important skills people possess, and our visual system provides particularly rich information in support of this skill. Yet, attempts to understand the human visual system, or to design an artificial solution for visual perception, have proven difficult. Human motion analysis has received much attention from researchers in the last decade due to its potential in a plethora of applications. This field of research focuses on the perception and recognition of human activities, including people identification. Human motion analysis is currently deployed in a wide range of applications such as visual surveillance in sensitive areas like airports and building lobbies. Moreover, the ability to analyse automatically the content of video sequences will be a potential key in the development of digital libraries for storing and retrieving data from video sequences using content-based queries [2]. This chapter is devoted to human motion perception and gait analysis. The systems employed for motion analysis along with the different stages involved are discussed. Lastly, the basic characteristics of human gait is described in more detail followed by a brief review of the current gait databases.

3.1 Automated Motion Analysis System

Motion capture systems for human motion analysis are dedicated to recording and measuring the posture variation of the human body over time [96]. The extraction of body parameters including positions and orientations with their temporal variation are used to reconstruct the three dimensional representation of human motion [68]. Such systems have received considerable interest due to their potential use in medical care, visual surveillance, gait analysis and virtual reality. Currently, there are two types of systems used to capture human motion: *marker-based* and *marker-less* systems. Marker-based solutions rely primarily on markers or sensors attached to key locations of the human body, whilst marker-less systems rely on computer vision techniques to extract motion features from the video stream. The majority of systems used for motion analysis are marker-based and are commercially available. This is mainly due to advance of technology and their perceived accuracy.

3.1.1 Marker-based Motion Analysis Systems

Extraction of data in marker-based systems is usually performed in two consecutive stages: sensing and processing [81]. In the first stage, the information about the joint positions and orientations are either received from sensors mounted on the human body, or extracted using special markers and high-speed pre-calibrated cameras. At the processing phase, the trajectory data of the joints are processed to reconstruct the 3D motion. Marker-based systems such as electro-mechanical, electro-magnetic and optical systems, have proven efficient and as a result are being deployed in a plethora of fields, particularly in clinical analysis and animation.

The *electro-mechanical motion capture systems* use an exoskeleton suit with sensors to capture the joints' angular motion [17]. Despite their accurate results, such systems require expensive hardware and are complex to set up. Furthermore, the mechanical suit might limit the human motion because of the cumbersome wiring around the human body and consequently leading to the measurement of unnatural motion pattern. Figure

(a) (b) (c)

FIGURE 3.1: Marker-Based Systems for Human Motion Analysis: (a) Electro-Mechanical System[1]. (b) Electro-Magnetic System[2]. (c) Optical system[3].

3.1(a) shows a subject wearing the Animazoo Gypsy5 electro-mechanical suit which costs approximately £12,000.

Compared to their *mechanical counterparts*, electro-magnetic systems use lighter suits and are more comfortable, therefore considerably increasing the range of human motion that is possible to capture. The electro-magnetic sensors are attached at key locations on the subject to record both the position and orientation of the sensors. Nevertheless, the wiring attached to the magnetic sensors may constrain the motion of the performer besides the difficulties encountered when changing the sensors' positions. As an example of the electro-magnetic capture system, the ShapeWrap II magnetic suit is depicted in Figure 3.1(b).

Optical systems use special markers such as reflective spheres attached at key locations of the human body such as the joints. Optical motion capture systems are widely used by the film industry as it is easy to setup and to change the markers' configuration compared to the electro-mechanical and electro-magnetic systems. In addition,

[1]http://www.animazoo.com
[2]http://www.measurand.com/products/ShapeWrap.html
[3]http://www.motionanalysis.com

the accuracy of the extraction is reported to be satisfactory for different types of motion such as running and jumping [96]. Figure 3.1(c) demonstrates an example of the use of optical motion capture in the film industry. The image was taken during the production of the "*I-Robot*" film. The actor shown in the figure is wearing a suit with reflective markers attached to it. The main drawback of using optical systems is their prohibitively expensive cost which is in the range of £50,000 to £150,000 [96]. Such systems require dedicated studios or laboratories with special lighting and multiple fast cameras to extract the markers, in addition to the licensing cost of the software used for the derivation of markers' trajectories. One of the major technical obstacles to optical systems lies in the difficulty to recover occluded markers, although this issue can be mitigated at the cost of an increased number of cameras, sometimes reaching 24 [44]. Another major challenge for optical systems is marker identification and correspondence, i.e, which marker is which. Unlike mechanical and magnetic systems where every sensor has a unique identifier, optical capture systems must determine which marker corresponds to a particular joint and establish a motion correspondence for markers across consecutive frames.

3.1.2 Marker-less Motion Analysis Systems

Alternatively, marker-less vision systems can extract the joint trajectories without placing special markers or sensors on the human body. In fact, all that is required is a set of video cameras with vision-based software. Marker-less motion capture systems are suited for applications where mounting sensors or markers on the subject is not an option, such as visual surveillance. Typically, the architecture of a marker-less system for human motion analysis consists of three main subsystems: i) detection and tracking of the subject, ii) feature extraction and iii) classification of human motion. These phases are explained in this section. Although this architecture is generic, the actual structure of such systems depends on their purposes, input types and information being processed. Figure 3.2 outlines the different subsystems involved in the process of an automated motion analysis system used for pedestrian detection and recognition. The same framework is being adopted in the course of this research.

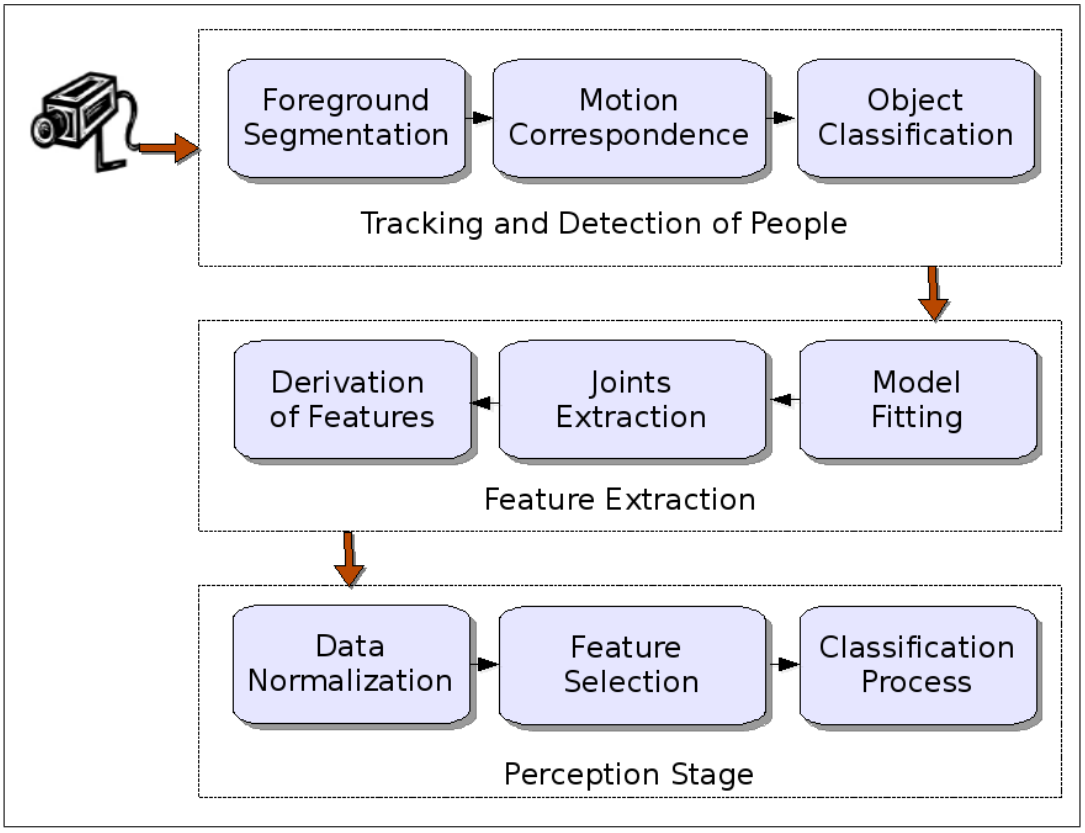


FIGURE 3.2: Structure of a Marker-less Motion Capture System.

3.1.2.1 Tracking of Moving Objects

The first step for an automated marker-less motion capture system is to detect moving objects such as people or vehicles in the scene. Usually, moving blobs are extracted by employing a variety of motion segmentation methods such as background subtraction [111], optical flow [80] and temporal differencing [77]. Having extracted moving objects in every frame from the video stream, tracking should be applied to establish a motion correspondence between detected moving blobs across frames. The tracking process involves matching moving objects in consecutive frames based on low-level features such as velocity, shape, colour and other visual information [4]. The criteria for selecting the best features are their robustness to noise, size changes and invariance to lighting and other environmental conditions. There is a trade-off between feature complexity and tracking efficiency. Low-level features such as points are easy to extract, but are more difficult to track than high-level features such as blobs. The classification of moving

objects and detection of people in the scene is performed either at the start or during the tracking process depending on the type of features used for the classification of moving obejcts. There are mainly two different types of features which are shape-based and motion-based.

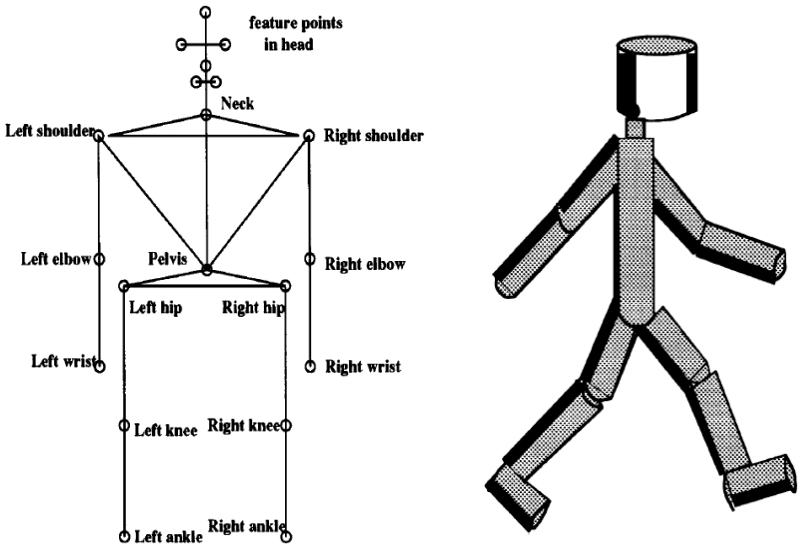
Feature correspondence is wherever possible established under well-defined constraints to eliminate most of the tracking problems [2]. Tracking methods are supported by prediction algorithms to estimate the parameters of moving objects in the next frame. This is based on motion models which describe how parameters change over time [81]. The most popular predictive method used for tracking is the Kalman filter [130], the Condensation algorithm [57], particle filtering [34] and the mean shift [24]. The Kalman filter which is based on Gaussian distribution for state estimation, is limited to situations where the probability distribution is unimodal. There are other derivatives of the Kalman filter such as the Extended Kalman Filter (EKF) [130] and the Unscented Kalman Filter (UKF) [65]. The condensation algorithm is proposed to support multimodel distributions by sampling the posterior distribution of the previous frame and propagating such samples to create the posterior of the current frame [57].

3.1.2.2 High-Level Feature Extraction

Feature extraction is the most important stage for automated marker-less capture systems since the crucial data required for the classification phase are derived at this stage. Feature extraction is the process of estimating the set of the measurements of high-level features in this case the configuration of the whole body or the configuration of the different body parts in a given scene and tracking them over a sequence of consecutive frames. High-level features estimated at this level for the perception phase include the joint angular measurements which are used for gait recognition [18], hands' configuration for gesture-driven applications [23] or silhouette-based features for pedestrian detection and recognition [41]. The methodologies used for feature extraction of human motion can be broadly classified into various categories depending on either the use of prior knowledge (model-based and model-free methods) or the type of features extracted

(pose recovery and average silhouette). In this review, we categorise the feature extraction methods proposed for human motion perception into three main classes namely: model-based, model-free and statistical-based methods.

In the *model-based approach*, an apriori model is used to match real images against a predefined model, and thereby extracting the features once the best match is obtained. The most commonly used model-based representations for feature extraction are the *stick* and *volumetric models* [139]. The stick figure model represents the human body as a combination of line segments that are connected at the joints. There are various ways to approximate the stick figure model such as the means of medial axis transform and distance transform [58]. Besides the stick model, the volumetric model is deployed to obtain a better geometric representation of the human shape at the expense of additional parameters. The volumetric model uses different 3-D shapes in its representation such as elliptical cylinders, cones and spheres. Figure 3.3 depicts both examples of the stick and volumetric models as examples of model-based methods.



(a) (b)

FIGURE 3.3: Model-Based Methods for Feature Extraction: (a) Stick Figure Model [20]. (b) Volumetric Model [46].

For the *model-free category*, no prior model is used for pose reconstruction, instead image features are mapped directly to trajectory data [81]. Feature estimation and

correspondence between successive frames are based on prediction, velocity, shape and texture. In [73], Kurakake *et al* worked on the extraction of joint locations by establishing correspondence between extracted blobs. The slight motion of objects between consecutive frames is the main assumption, whereby feature correspondence is conducted using various geometric constraints.

In the *statistical-based approach*, most of the methods are either silhouette-based or contour-based [82]. The human motion is usually perceived without the need to extract the body parts or joints. Instead, motion features are derived directly from video frames by applying statistical methods on the silhouette or contour data. Wang et al [127] derived binary silhouettes of walking subjects which are converted into a one-dimension normalised distance signal by contour unwrapping with respect to the centroid position. In the same way, Huang [56] and Foster [41] dealt with the problem of gait recognition using human silhouettes to derive gait signatures for walking subjects. They applied statistical methods on silhouette data including Principal Component Analysis (PCA) and Fisher Discriminant Analysis.

3.1.2.3 Classification Phase

The perception process is mainly a pattern classification problem which involves matching a test sequence with an unknown label against a group of labelled references considered as the training set [43]. At this stage, a high-level description is produced from the features extracted from previous phases. This description includes activity or gesture recognition, gender classification, medical assessment or gait recognition. The classification process is normally preceded by pre-processing stages such as data normalisation, feature selection and dimensionality reduction of the feature space. A variety of pattern recognition methods is being used in vision-based systems for the perception of human motion, including Neural Networks, Support Vector Machines (SVM) and K-Nearest Neighbour method (KNN).

3.2 Human Motion Perception

Although people can discern the state of the subject from a single static image, motion pictures provide even richer and reliable information for the perception of the different biological, social and psychological characteristics of the person [15] such as emotions, actions and personality traits of the subject. Furthermore, this notion was also observed by Darwin (1872) in his book "*The Expression of Emotions in Man and Animals*" where it was stated:

> "*Actions speak louder than pictures when it comes to understanding what others are doing.*"

The human visual system is very sensitive to motion as it tends to focus attention on moving objects. In contrast, static or motionless objects are not as straightforward to detect. Motion is a spatio-temporal event defined as the change of spatial location over time. Given some visual input, the visual perception of motion is regarded as the process by which the visual system acquires perceptual knowledge such as the speed and direction of the moving object [33]. Whilst this process is spontaneous for the human visual system, it has proven to be extraordinarily difficult to duplicate this capability into computer vision systems.

Psychological studies carried out by the Swedish psychologist Johansson [63] in 1973, revealed that people are able to perceive human motion from Moving Lights Display (MLD). An MLD is a two-dimensional video of a collection of bright dots attached to the human body taken against a dark background where only the bright dots are visible in the scene. An observer can recognise different types of human motion such as walking, jumping, dancing and so on. Moreover, the observer can make a judgement about the gender of the performer [70], and even further identify the person if they are already familiar with their gait [45]. Although the different parts of the human body are not seen in the points and no links exist between the bright dots to show the skeleton structure of the human body, the observer can recover the full structure of the moving object. Thereby, the motion of the joints contains sufficient information for the perception of

human motion [14, 36]. Figure (3.4) shows a series of frames used for the MLD. The joint positions of walking subjects are shown in every frame.

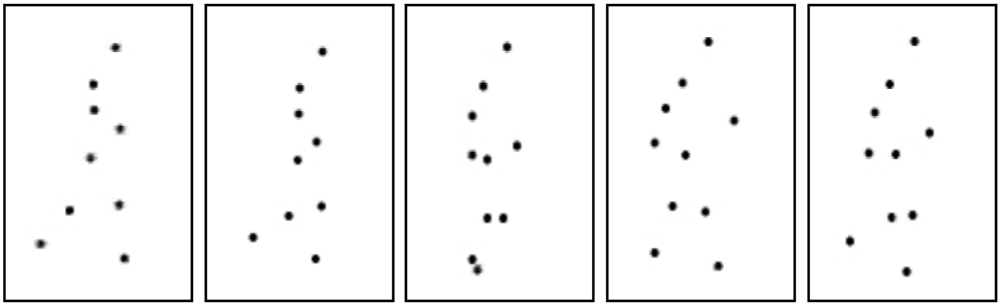


FIGURE 3.4: Frames taken from a Moving Light Displays video.

There is a wealth of research which strives to document the capability of the human visual system to perceive the human motion from a small number of moving point. Nevertheless, the underlying perceptual process is poorly understood and there is still a lack of research which explains the underlying principles for representing and retrieving the biological motion [117]. Two main theories have been put forward for the perception of human motion from the MLD: *structure-based* and *motion-based* [21]. The former theory claims that the initial step is recovering the 3D structure from the motion information in MLDs, and then use the recovered structure for the purpose of recognition. In the latter approach, recognition is based directly on the motion information without recovering the skeleton structure of the human body from the MLD; instead the motion information is extracted from a sequence of frames.

The motion of the human body is a form of non-rigid and articulated motion [3]. Hence, detecting and tracking is a very difficult task as the non-rigid motion encompasses a wide range of possible motion transformations due to the highly flexible structure and the opportunity for self-occlusion [139]. On the other hand, during walking and running, people share the same global gait pattern, as they swing their legs the in order to move. This constrains the allowable motion transformation. Therefore, gait motion can be considered as an ideal starting point for motion analysis due to its global and cyclic nature.

3.3 Gait for Human Motion Analysis

3.3.1 Gait Analysis

Gait analysis is the systematic study of human walking [131] aimed at the quantification and understanding of the locomotion process. This study involves the observation of body movements, mechanics and muscle activities. Gait analysis is carried out for two main purposes [131]: firstly, the treatment of patients with gait abnormalities, secondly to enhance the knowledge and understanding of human gait. The study of human gait dates back to the ancient times. Aristotle (384-322 BC) might be considered the first person to study and describe gait in his book "*De Motu Animalium*" (on the movement/motion of animals). Aristotle wrote about the difference between human and animals, and his observations of human gait were concluded as follows:

> "*If a man were to walk on the ground alongside a wall with a reed dipped in ink attached to his head, the line traced by the reed would not be straight but zigzag, because it goes lower when he bends and higher when he stands upright*"

Leonardo da Vinci (1452-1519) described the principles of walking and observed the complexity and symmetric nature of human gait in his famous anatomic paintings. During the 18th century, the Weber brothers in Germany conducted the first formal biomechanical experiment, giving clear description about the timing of gait cycle. In 1892, Muybridge devised an apparatus with multiple trip wires attached to the camera shutters which was employed to record the locomotion process [86]. Gait analysis evolved greatly in the 20th century with the invention of many tools and instruments needed for the measurement and quantification of gait motion [92].

3.3.2 Human Gait Characteristics

Gait is defined as the manner of locomotion characterised by consecutive periods of loading and unloading the limbs. Gait includes running, walking and hopping. However the

term gait is most frequently used to describe walking. The rhythmic pattern of human gait is performed in a repeatable and characteristic way [132]. The locomotion process involves the interaction of many body systems working together to yield the most efficient walking pattern [40]. The locomotion system consists of four main subtasks that are fulfilled at the same time to produce the walking pattern [135]. These four functions are: (i) initiation and termination of locomotor movements (ii) the generation of continuous movement to progress toward a destination (iii) adaptability to meet any changes in the environment or other concurrent tasks (iv) maintenance of the equilibrium during progression. Compared with quadrupeds, the maintenance of stability and balance for humans during walking is a particularly difficult task for the postural control system. This is mainly because for most of the gait cycle, the human body is supported by only a single leg with the centre of mass passing outside the base of support provided by the foot in contact with the floor [134].

Early medical investigations conducted by Murray *et al.* [85] produced a standard gait pattern for normal walking people aimed at studying the gait patterns for pathologically abnormal patients. The experiments were performed on sixty people aged between 20 and 65 years old. Each subject was instructed to walk for a repeated number of trials. For the collection of gait data, special markers were attached on every subject. Murray suggested that gait is unique for every subject if all gait movements are considered. It was reported that the motion patterns of the pelvic and thorax regions are highly variable from one subject to another. However, the extraction of such patterns is complex using computer vision methods. In [84], Murray observed that the ankle rotation, pelvic motion and spatial displacements of the trunk embed the subject individuality due to their consistency at different trials. Although, there is a wealth of gait studies in the literature aimed for medical use, none is concerned for the use of gait for biometrics and recognizing people. The gait measurements and results introduced by Murray prove to be of benefit to the use of gait recognition using computer vision methods to be shown later.

3.3.3 Human Gait Cycle

A gait cycle is defined as the time interval between successive instances of initial foot-to-floor contact for the same foot [27], and the way a human walks is marked by the movement of each leg. Each one possesses two distinct phases. When the foot is in contact with the floor the leg is at the *stance phase*. The time when the foot is off the floor to the next step is defined as the *swing phase*. Each phase is marked by a start and an end; the stance phase begins with the *heel strike* of one foot when the leg strikes the ground.

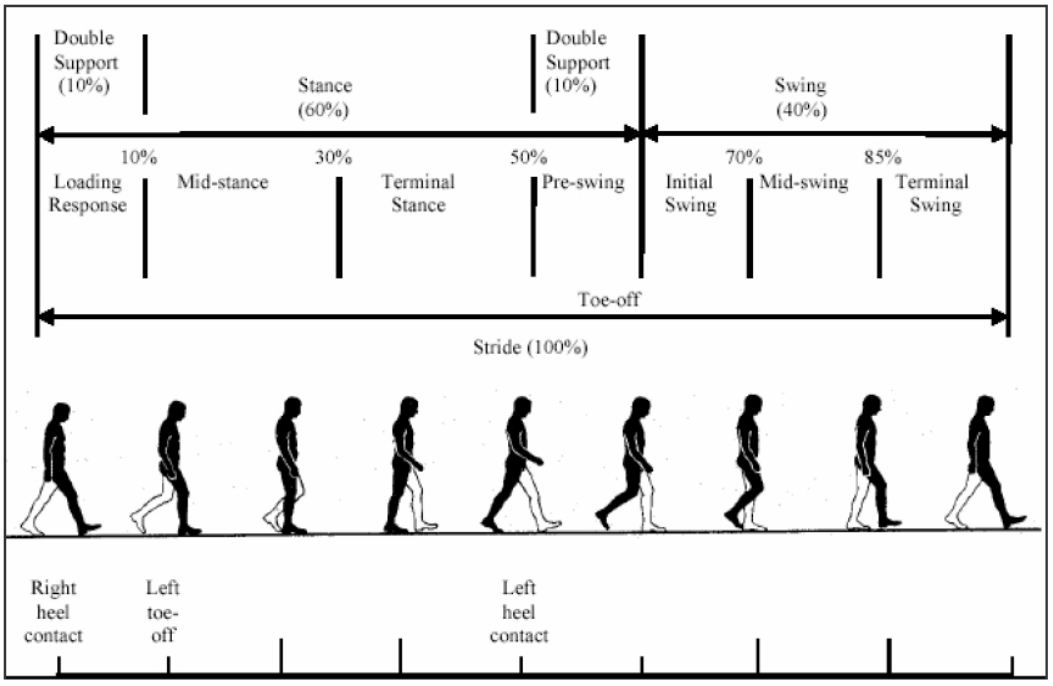


FIGURE 3.5: Human Gait cycle [131].

Considering the motion of the left foot, its heel strike starts its stance phase, which is characterised by a sequence of events. To bring the left foot onto the ground the ankle flexes, and as a result, the body weight is transferred onto it. The right leg then swings through in front of the left leg as the left heel lifts off the ground; this is referred to "heel off". As the body weight moves onto the right foot, the supporting left knee flexes, the stance phase ends, when the remainder of the left foot, which is now behind, lifts off the ground. This is referred to as "toe-off" and it occurs before the swing phase. As a result

of the toe off, the weight is transferred onto the right leg and the left leg swings forward to strike the ground in front of the right foot. The swing phase ends with the heel strike of the left foot. A stance and a swing phase form a cycle referred to as the gait cycle which is illustrated in Figure (3.5). Stride length is considered the linear distance in the plane of progression between successive points of contact of the same foot. Step length is the distance between successive contact points of opposite feet. A step is the motion between successive heel strikes of opposite feet [27].

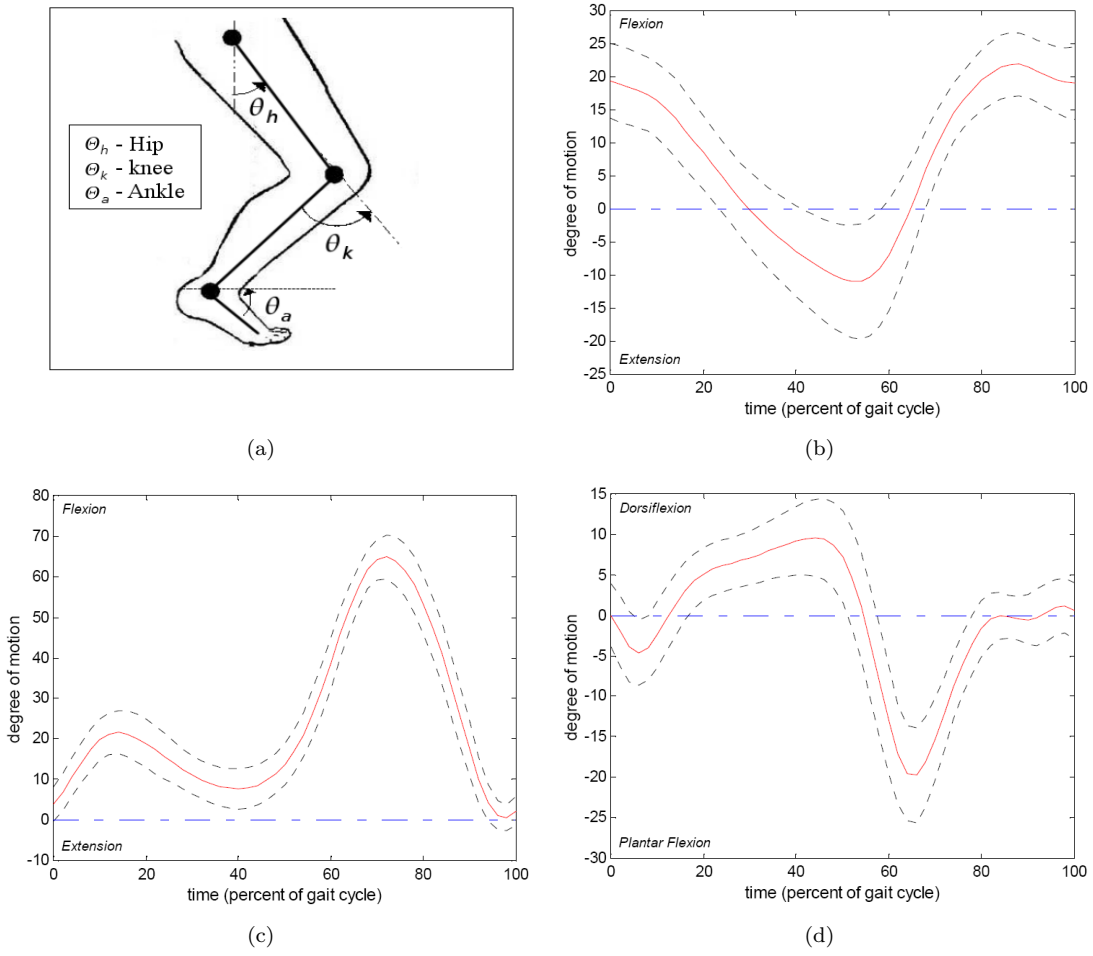


FIGURE 3.6: Gait Angular Motion: (a) Gait Angles. (b) Hip Angular Motion. (c) Knee Angular Motion. (d) Ankle Angular Motion.

3.3.4 Gait Angular kinematics

The gait angular kinematics of the hip, knee and ankle angles are illustrated in Figure (3.6). The hip initially bends or flexes by approximately 20° throughout the terminal stance phase, then it extends until it reaches approximately 10 degrees during the stance phase. During the pre-swing and throughout most of swing phase, the hip flexes to nearly 20° , and then starts to extend just before the next initial contact. The knee angular motion illustrated in Figure (3.6(c)), shows the knee is almost fully extended then during the first part of the midstance, it gradually begins to flex to its support phase peak which is about 20 degrees. The knee extends again almost fully and then flexes to approximately 40° during the pre-swing phase. After toe-off, the knee flexes to reach a peak of 60 to 70 degrees (measured relative to the thigh) at mid-swing, then extends again in preparation for the next initial contact. The angular pattern of the ankle described in Figure (3.6(d)) shows that the ankle extends to about a maximum of 7° (measured relative to the horizontal plane) from an initial angle. During the mid-stance phase, it flexes to a maximum of 15 degrees as the lower leg rotates over the supporting foot. Then the ankle extends to approximately 20° during the terminal stance and pre-swing. After toe-off , the ankle rapidly flexes to the start position.

3.3.5 Analysis of Joint Spatial motion

To analyse the spatial displacement of the joint motion, the joint positions from 30 video sequences with people walking normal to the viewing plane of the camera, have been manually labelled. The videos were taken from the SOTON database described in Section (3.4.1) recorded at frame rate of 30 frames/seconds. For every frame of the video sequence, the positions of the right ankle, right knee and hip were manually labelled. The data for the ankle between two consecutive heel strikes of the same leg are normalized between the range $[0, 1]$ as shown in Figure 3.7(a). Whilst, we have normalized the data for the knee and hip extracted between two consecutive stances of the same leg as shown in Figures 3.7(c) and 3.7(e). The corresponding horizontal displacement for each joint is plotted against the motion graph of the joint in Figure (3.7).

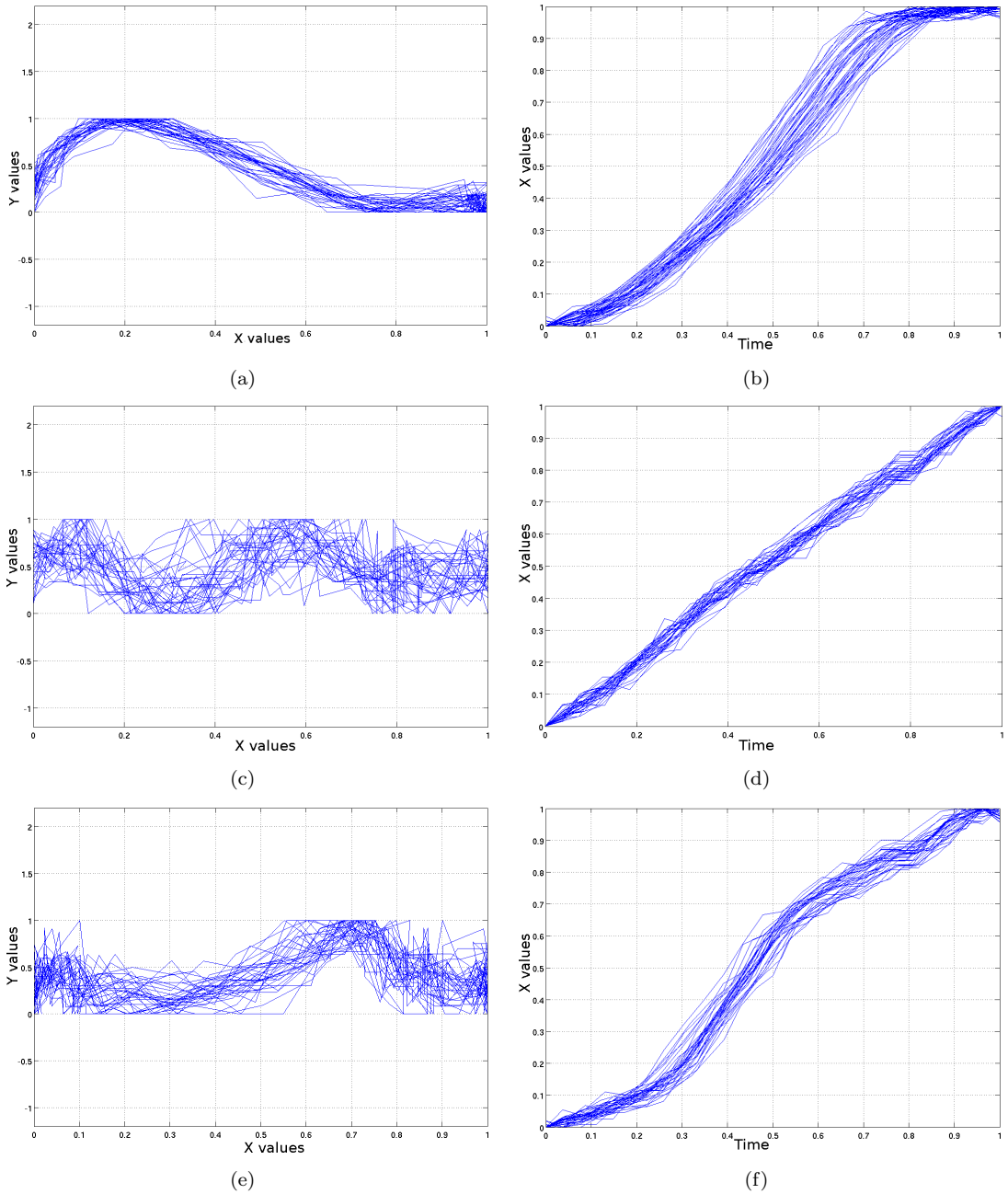


FIGURE 3.7: Spatial Motion Analysis of the Joints:(a), (c), (e) Angular Motion Graphs for Ankle, Hip, Knee respectively. (b), (d), (f) Horizontal Displacement for Ankle, Hip, Knee respectively.

It can be observed that people have more or less the same ankle motion pattern. Another graph 3.7(b) is plotted showing the horizontal displacement of the ankle, where it is noted that the graphs for all subjects nearly coincide, leading to the suggestion that for a normalized data set, subjects move their ankles forward with the same velocity. Figures 3.7(c) and 3.7(e) show the hip and knee motions respectively for the normalized extracted data. In contrast to the smooth graphs of the ankle, there is noise in the data for the hip and knee due the difficulties encountered during the manual labelling. Nevertheless, it can be observed that walking people largely share the same global pattern for the hip and knee motions. The horizontal displacement for the hip and knee are shown in Figures 3.7(d) and 3.7(f). The hip forward velocity is approximated to be constant for all subjects, in contrast to the knee velocity, which varies for all subjects. However, with the data normalized, people have more or less the same knee horizontal velocity.

3.4 Gait Databases

As gait analysis has gained an increasing interest in different research areas, along with the recent scientific developments of using gait as a potential biometric, the establishment of gait databases has become vital for the evaluation and assessment of research theories and systems proposed for gait analysis, automated marker-less extraction and gait recognition. There were two early gait databases which are the UCSD and Southampton database. The first database was collected by the Visual Computing Group at the University of California San Diego. There are 6 subjects in the database filmed outdoors. The Southampton database is recorded indoors and consists of 16 video sequences for 4 subjects wearing special trousers [92].

Recently, several gait databases were developed primarily for the HumanID at a Distance program [98] sponsored by the Defence Advanced Research Projects Agency (DARPA). The program was aimed to improve technologies for facial and gait recognition as well as new technologies for people identification. The HumanID project included the following research institutions: University of Southampton [107], University of Maryland, Georgia Institute of Technology and Massachusetts Institute of Technology. In this

research, we used the SOTON gait database developed by the Information: Signals, Images, Systems (ISIS) Research Group at the University of Southampton [107] for the analysis and evaluation of automated extraction of gait features as well as gait recognition. Table (3.1) surveys the different gait databases developed under the DARPA program.

Database	Sub[1]	Seq[2]	Scene Description	Covariate Factors
Covariate SOTON	12	12,730	Indoor conditions	Footwear, clothing, walking speed, viewpoint and carrying conditions
Large SOTON	118	10,442	Indoor, Outdoor and Treadmill	Viewpoint
Gait Challenge	122	1,870	Outdoor environment	Viewpoint, surface, footwear, time and carrying conditions
CMU	25	600	Indoor and Treadmill	Walking speed, viewpoint, surface and carrying conditions
GATECH	15	168	Outdoor	Viewpoint
MIT	24	194	Indoors	

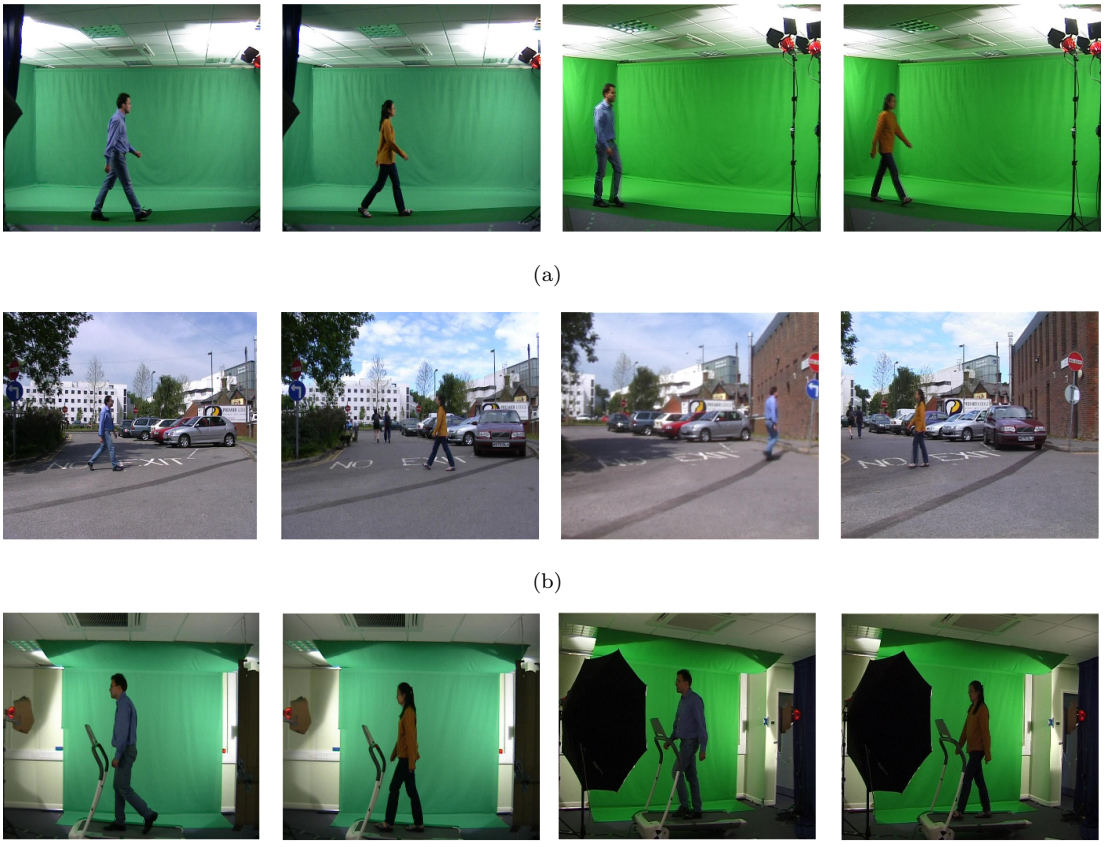
TABLE 3.1: Human Gait Databases.

3.4.1 Southampton (Inter-Subjects) Large database

The SOTON database is the largest database [19] containing more than 10,000 video sequences of walking subjects. The database is aimed to investigate the potential of gait as a biometric as well as to assess the capabilities of computer vision methods for automated extraction of gait features. There are 118 different subjects in the database with more than 40 video sequences for every person, henceforth it is considered the largest database for the analysis of between-subject variation [92]. Subjects were filmed in indoor and outdoor environments using high quality progressive and interlaced digital video at resolution of 720 by 576 pixels. For the indoor data, people walked on a special track inside a laboratory with controlled lighting and chroma-key background.

For the data recorded outdoors, there is no control over the background and lighting conditions. Subjects were filmed walking in both directions for indoors and outdoors. To account for viewpoint invariance of gait analysis, two digital camcorders are used

for the recording of video sequences in indoor and outdoor environments. In order to capture the sagittal view, one camera was placed normal to the walking direction whilst the other camera was set to record from an oblique angle. The video cameras used for sagittal and oblique views were set to progressive scan and interlaced respectively. Figure (3.8(a)) depicts samples of the indoor database of two subjects recorded from sagittal and oblique views. Examples of the outdoor database are shown in Figures (3.8(b)). In addition to the indoor and outdoor data, the SOTON gait database comprises treadmill data as depicted in Figure (3.8(c)). Subjects are recorded walking on the treadmill at constant speed with similar laboratory conditions as the indoor data.



(a)

(b)

(c)

FIGURE 3.8: The Southampton Large Gait Database: (a) Indoor Sagittal and Oblique View. (b) Outdoor Sagittal and Oblique View. (c) Treadmill Data.

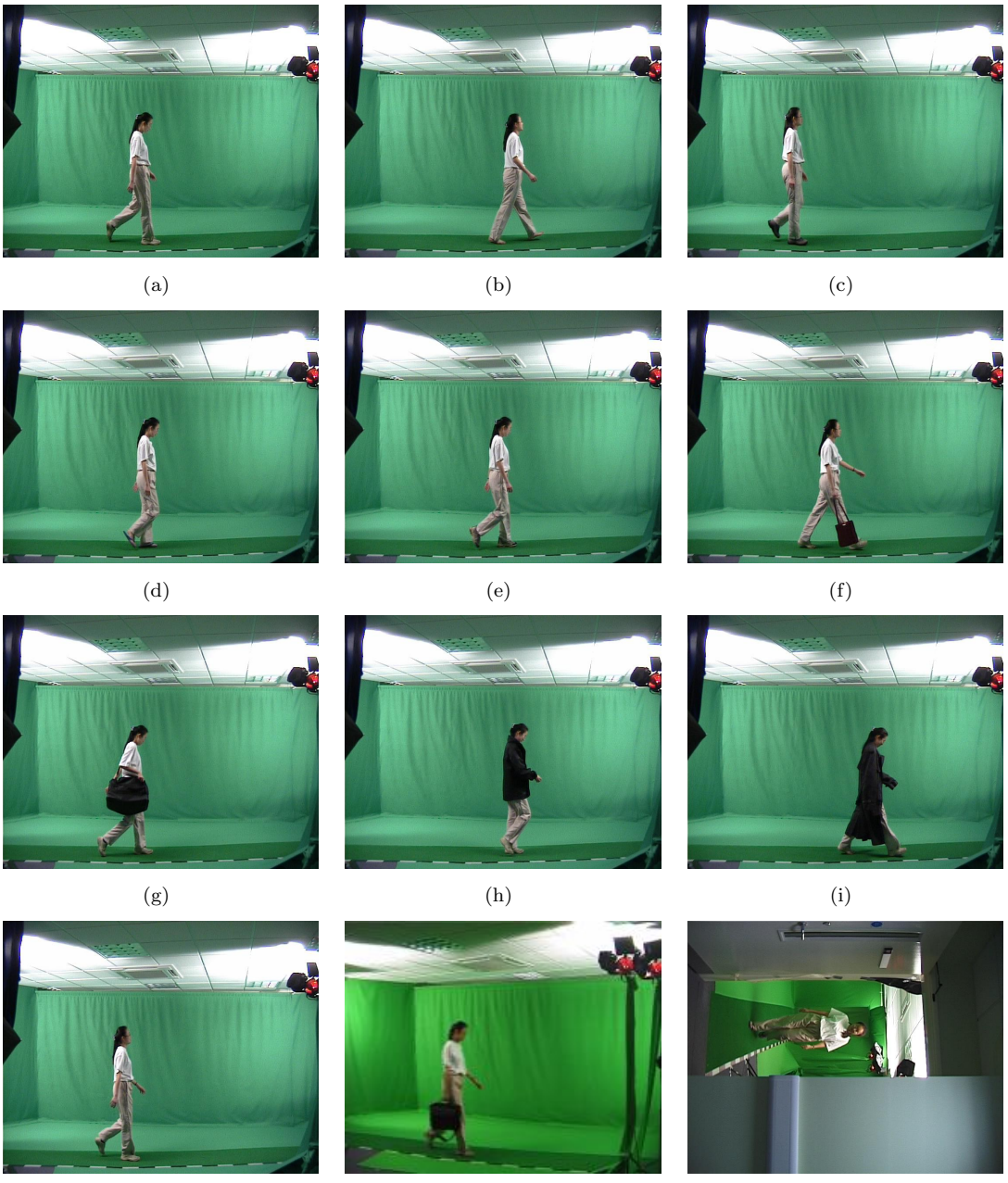
[1]Number of subjects
[2]The total number of sequences in the database

3.4.2 Southampton Covariate Database

The SOTON covariate database is constructed for the purpose of examining and assessing the different covariate factors which may affect gait analysis and recognition. The database consists of 12,730 video sequences for 12 different subjects who were also filmed in the SOTON large database. Subjects were recorded walking on a special track inside a laboratory with the same indoor conditions described earlier for the SOTON large database. Each subject was recorded walking in 15 different scenarios with at least 10 trials for each case. The covariate database includes cases where subjects wear a variety of footwear (flip flop, boots, trainer, socks, shoes), cloths (trench coat, rain coat, normal clothing) and carrying various bags (hand bag, barrel bag). Besides, people were filmed walking at different speeds (normal, slow and quick). Four different viewpoints were used for the database including sagittal, frontal, and two oblique views. Figure (3.9) shows the different covariate factors of the database.

3.5 Conclusions

Human motion analysis is one of the most active and attractive research areas in computer vision due to the wide vistas that it opens for a plethora of fields ranging from visual surveillance to medical assessment. Studies conducted by psychologists revealed that joints' trajectories possess sufficient information to recognise the subject identity. Nevertheless, the extraction of motion features has proved to be a difficult task, to say the least, due to the complex nature of human motion. As the human gait pattern is global and periodic, such rhythmic motion is considered as an ideal and attractive starting point for people detection and recognition. The main thrust behind the recent development of gait as a recognition method, is the complementary studies from psychology and various other disciplines which supported the founding concept that gait is unique for every person and people can recognise each other by the way they walk. This information will be used in our model-based analysis. The evaluation of proposed methods will use the data that has been described here.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

FIGURE 3.9: The Southampton Covariate Gait Database: (a) Normal Walking. (b) Barefeet. (c) Boots. (d) Flip Flops. (e) Trainer. (f) Hand Bag. (g) Barrel Bag. (h) Coat. (i) Trench Coat. (j) Quick Walking. (k) Oblique View. (l) Frontal View.

Chapter 4

Gait for People Detection

As discussed in the previous chapter, the first stage in an automated visual surveillance system is the tracking and detection of subjects from the video stream. Tracking and classifying moving objects is of prime importance for most applications dealing with human motion. This is because it lays the foundation for the subsequent stages to extract and derive features required for the perception phase. As most of the methods proposed for people detection are largely motion-based, the use of gait for pedestrian detection from real-world video surveillance is investigated in this research as a first step for a biometric system. In this chapter, the proposed method for tracking multiple moving regions in an unconstrained environment is described. The rhythmic pattern of gait motion is utilised as the main cue to distinguish walking subjects from other moving objects. The experimental results for tracking and classification of moving objects are drawn at the end of the chapter.

4.1 Foreground Segmentation

4.1.1 Background Subtraction

The first stage for an automated surveillance system is the detection of moving objects from the scene. This is often performed via background subtraction. Moving objects are detected by taking the difference between the current image and the background image in a pixel-by-pixel fashion for the case when the background is static. The approach employed for the segmentation of moving objects in this study is the adaptive background subtraction proposed by Stauffer and Grimson [111]. A mixture of K (from 3 to 5) Gaussian distributions is used to model the RGB colour changes. The probability of a pixel to have intensity x_t at time t is given by:

$$P(x_t) = \sum_{j=1}^{K} w_j \eta\left(x_t, \mu_j, \Sigma_j\right) \tag{4.1}$$

where w_j, μ_j and Σ_j are the weights, mean, and covariance matrix for the j^{th} distribution respectively. η is the normal density function defined in (4.2):

$$\eta(x, \mu_j, \Sigma_j) = \frac{1}{(2\pi)^{d/2}\left|\Sigma_j\right|^{1/2}} e^{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)} \tag{4.2}$$

d is the dimension of the colour model which is 3 for the RGB model. The distributions are ranked according to the ratio of their weights over standard deviations. The background model is formed using the first B distributions such that B is estimated using (4.3):

$$B = \arg\min_{b} \left(\sum_{j=1}^{b} w_j > T \right) \tag{4.3}$$

where $T \in [0, 1]$ is a threshold that defines the number of modes of variations in the background. A small value of T would result a strict foreground segmentation with only uni-modal background surfaces are considered and vice versa for larger values of T. Background subtraction is performed by labelling any pixel that is more than 2.5

standard deviations of any of the B distributions as a foreground pixel. The parameters of the distribution μ_j and Σ_j are updated recursively [111].

4.1.2 Shadow and Noise Suppression

Since the adaptive background subtraction lacks capability to remove shadows, we used the approach described by Cucchiara *et al* [25] to evaluate whether a foreground pixel corresponds to cast shadow based on the Hue Saturation Value (HSV) colour information. The chromaticity and luminosity of the foreground pixels are separated using the HSV colour space which is proved to match the human perception of colour more closely than the RGB model [52]. The proposed method [25] assumes that shadows reduce surface brightness and saturation while maintaining chromaticity properties in the HSV colour space. In [25], shadow mask is defined for every foreground segmented pixel (x, y) with the following conditions.

$$SP_k(x,y) = \begin{cases} 1 & if \alpha \leq \frac{I_k^v(x,y)}{B_k^v(x,y)} \leq \beta \\ & \wedge (I_k^s(x,y) - B_k^s(x,y)) \leq \tau^s \\ & \wedge (I_k^h(x,y) - B_k^h(x,y)) \leq \tau^h \\ 0 & otherwise \end{cases} \tag{4.4}$$

such that $I_k^h(x,y)$, $I_k^s(x,y)$ and $I_k^v(x,y)$ are the intensity values for the three HSV components at coordinate (x, y) in the frame k. B is the background image. The α parameters defines how the strong the light source, whereas β is set to avoid detection of pixels where background changed slightly by noise as shadows. The choice for the threshold parameters τ^s and τ^h is set experimentally to 0.1 and 0.5, respectively.

Figure (4.1) shows the results of adaptive background subtraction followed by shadow suppression. Morphological operators including erosion and dilution are employed to remove noise from the foreground image produced from the background subtraction process. Finally, Connected Component Analysis is utilised to extract the different blobs from the foreground binary silhouettes.

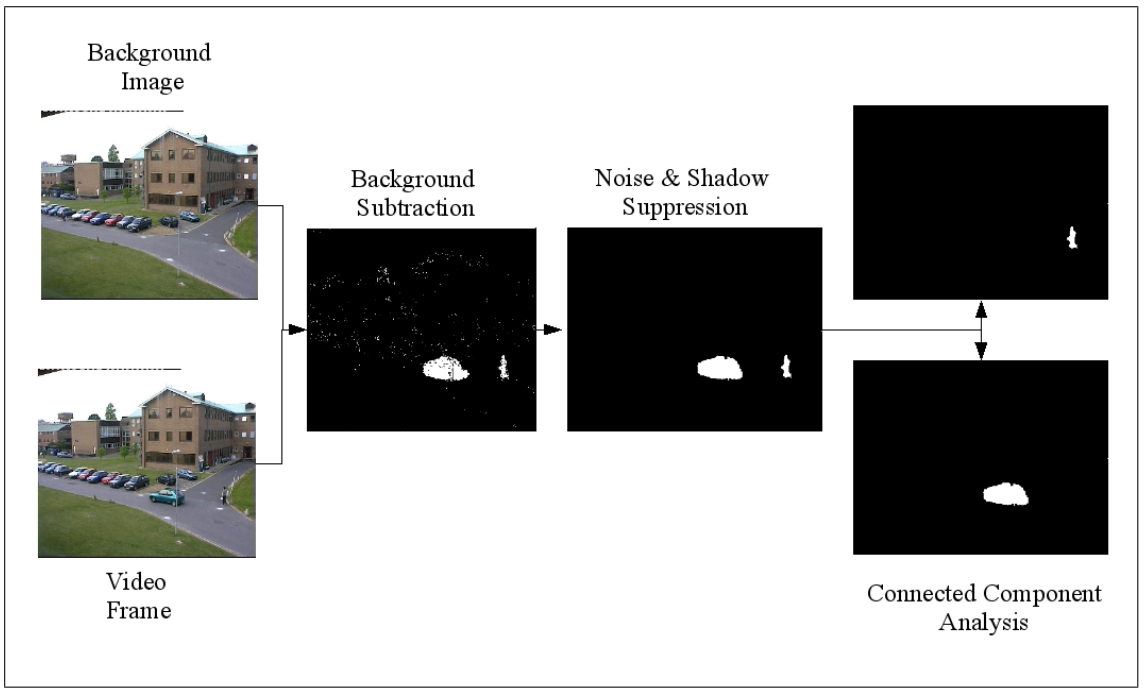


FIGURE 4.1: Foreground Segmentation

4.2 Tracking Moving Objects

Having detected moving regions using the foreground segmentation algorithm described in the earlier section, the next step is to track moving objects over a sequence of frames. Tracking multiple objects simultaneously is a fundamental component in surveillance applications and not surprisingly a challenging task as well. During this process, typical complications encountered in real world surveillance video need to be meticulously considered such as occlusion, shape and lighting variations, background clutter (such as moving tree leaves.) and overlap of moving objects; for which a robust region correspondence algorithm is required. Tracking systems must be able to track objects which are partially occluded and capable of recovering objects which were fully occluded for a short time. Furthermore, the entry of new objects into the scene must be handled correctly such as when a new walking subject enters the scene, or an existing object splits from moving region as for the case when a subject leaves a group of people walking together. Moreover, the exit of moving objects must be taken into consideration along with the

case where moving objects become stationary (e.g. a moving vehicle stops) as well as the situation where separate moving objects merge into one moving region.

Existing tracking methods use various cues such as size, compactness and colour intensity. The proposed approach in this research models moving objects as temporal templates characterised by a combination of three basic features namely: the size, the centroid position, and the aspect ratio of height to width of the object bounding box. Furthermore, the colour information could be used for tracking by taking the major or dominant colour components derived using statistical methods such as Principal Component Analysis (PCA). In this study, only the shape-based features are considered because they involve low-complexity computation and yet they enjoy robust characteristics which are highly sought after in the design and implementation of a tracking system suitable for real-time applications. The size, which is a shape-based feature that provides information about the shape of the moving region, is measured as the number of pixels contained within the object. In general, moving objects tend to keep a similar size to the one in the previous frame. Since moving objects move slightly from frame to frame, the centroid position is considered a strong feature for tracking. The centroid position is computed as the average position of all pixels within a region R of N pixels as expressed by the following equation:

$$(\bar{x}, \bar{y}) = \left(\frac{1}{N} \sum_{P_{(x,y)} \in R} x, \frac{1}{N} \sum_{P_{(x,y)} \in R} y \right) \tag{4.5}$$

Furthermore, combining features would overcome the challenges encountered during the tracking phase such as shape variation, lighting and segmentation errors. A key element to successful tracking is to maintain the state of moving objects over the sequence of frames consistently. For this reason, moving objects are assigned to different virtual *layers* such that moving regions which correspond to the same object are allocated to the same layer. The *layer* can be viewed as a virtual container proposed to ease the tracking process. Layers are defined by the following parameters $L_i < s_i, h_i, w_i, x_i, y_i >$ where i is the layer index. s_i, w_i and h_i are the values of the size, width and height parameters for objects belonging to the i^{th} layer. x_i and y_i are the predicted centroid

position of the object in the next frame. The computation method of these variables is described in the next section.

4.2.1 Region Correspondence Algorithm

At the initial frame, detected moving objects are allocated to new layers whereby we update the parameters for every layer. In the following frames, we create a list which contains all the existing active layers. As most blobs with smaller size are noise and the size is an important cue for tracking, newly detected moving regions are ordered according to their size. Then, starting from the larger blobs we take every object and search for the ideal layer in the list to assign this object to. The allocation criteria is discussed in the following sections. If an allocation is made, the parameters for the selected layer are updated accordingly and we remove this layer from the list. If an object is not assigned to one of the existing layers, a new layer is created for this new object. It is possible that some active layers remain in the list if no corresponding candidate objects have been allocated to them. This may occur in two cases: either because of full occlusion or the total disappearance of the object from the scene. In such situations, the centroid parameters for these layers are updated based on the priori knowledge from the previous frames.

In order to maintain the state of moving objects consistently during the tracking phase, the layer parameters are updated using different update functions. As objects move relatively slowly with respect to the video frame rate (which is usually 25 to 30 frames/second for the set of video data being used for pedestrian detection and gait analysis), the centroid position is estimated linearly by computing the velocity V_i as the spatial difference of the last two previous positions. The Kalman filter method can also be employed to predict the centroid position. The size, width and height variables are updated using a linear model constructed in such a way that new information is added slowly and old information is gradually forgotten. This allows the model to gradually accommodate the changes of the object shape. For a feature f_t of a layer with a new candidate object whose measurement is m_t at the t^{th} frame, the new feature f_{t+1} at

$(t+1)^{th}$ frame is updated using the function defined in (4.6) where ε is a small number typically set to 0.1.

$$f_{t+1} = \frac{f_t + (m_t \times \varepsilon)}{1 + \varepsilon} \tag{4.6}$$

Because the Euclidean distance metric allows dimensions with larger scales and variances to dominate the feature space, the cost function C for allocating moving objects to their corresponding layers is based on the Mahalanobis metric measure using Equation (4.7). The use of the Mahalanobis metric alleviates most of the Euclidean metric limitations, as it accounts automatically for the scaling of coordinate axes in the feature space [140, 129]:

$$C = \sqrt{(\mathbf{v}_l - \mathbf{v}_c)^T \Sigma^{-1} (\mathbf{v}_l - \mathbf{v}_c)} \tag{4.7}$$

where $\mathbf{v}_l$ and $\mathbf{v}_c$ are the feature vectors for the layer and the candidate object respectively. Σ^{-1} is the inverse of the within-class covariance matrix of the training set [129] . The proposed algorithm used for the tracking of moving objects is described in Figure (4.2). The method employs a number of constraints and conditions which are explained in the following sections.

4.2.2 Tracking Constraints

A number of constraints are imposed on the allocation criteria in order to handle occlusion, entry and exit of moving objects into the scene as well as the split and merge of moving regions. A candidate object will be allocated to layer L_i only if (1) the layer L_i has the smallest distance C computed using the Mahalanobis metric measure and (2) the size of the object is within the sizes of the objects allocated to this layer which is expressed statistically as $|s_i - S| < 2.5\sigma_i$ where S is the size of the detected moving object, σ_i and s_i are the standard deviation and mean values of objects' sizes belonging to the i^{th} layer. To cope with the appearance of uninteresting regions and noise such as background clutter, two threshold values S and F are introduced. The threshold S is a size filter such that regions with a size less than this threshold are considered noise and

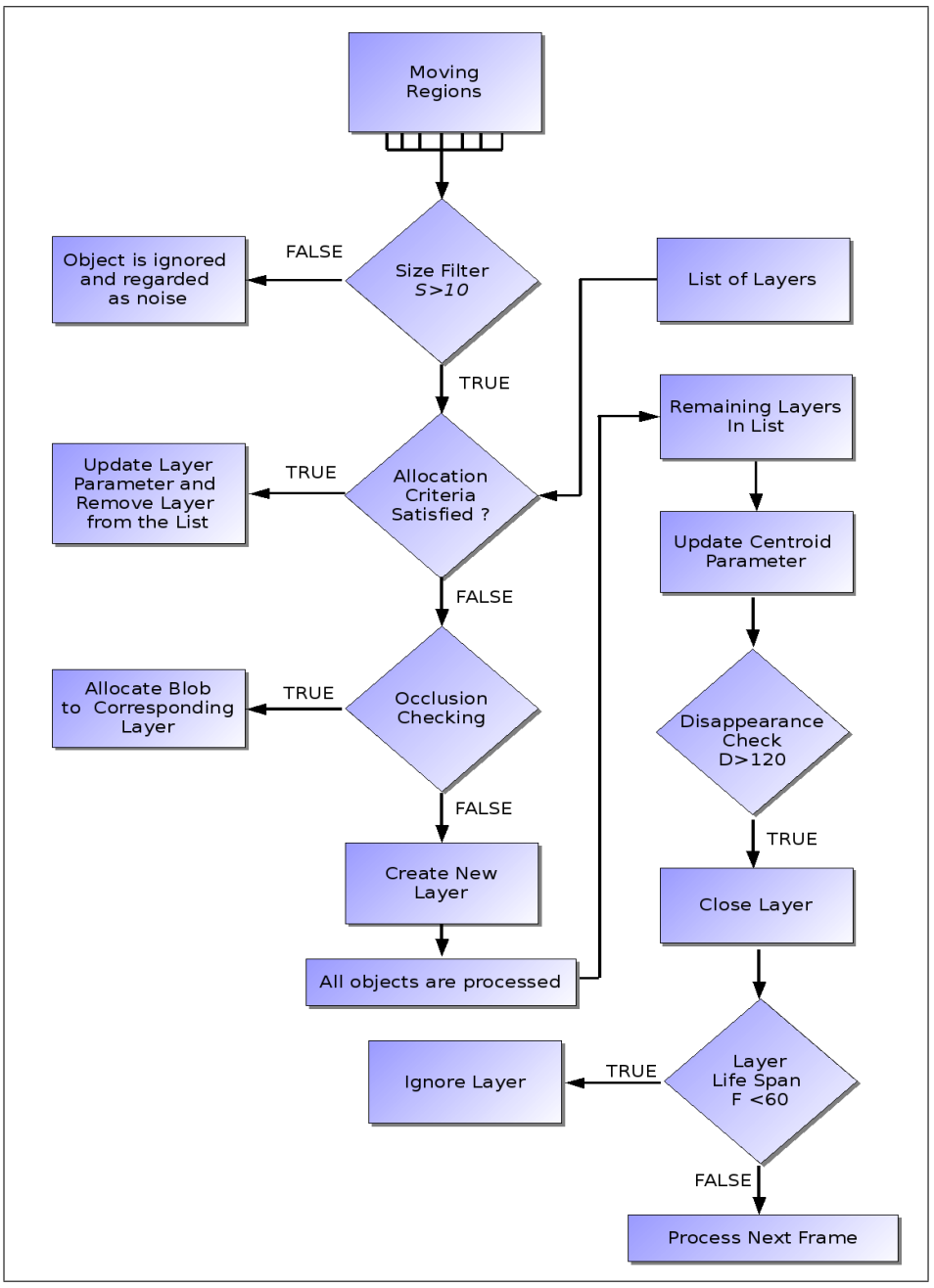


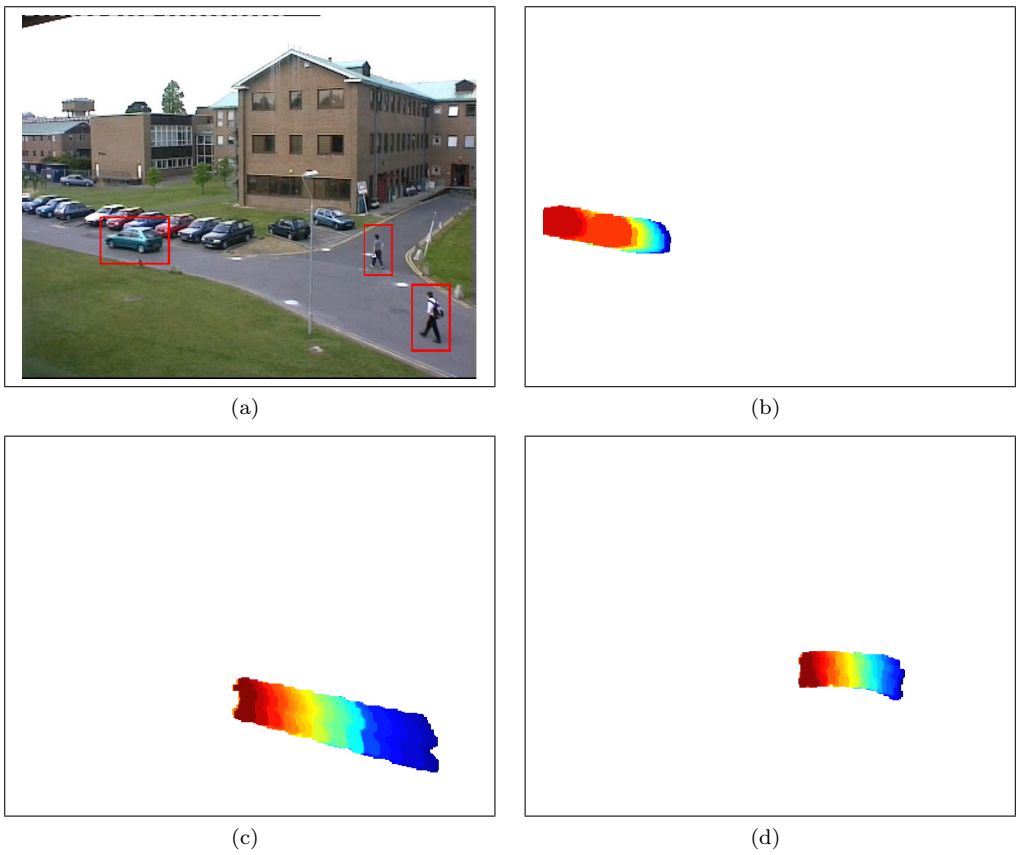
FIGURE 4.2: The Block Diagram for the Tracking Algorithm. D is the number of consecutive frames such that a layer does not have a corresponding moving object.

therefore removed. The constant F is a life span threshold such that if a layer has a life span of F frames or less, then this layer is ignored and deleted. The values of S and F depend on the resolution and the frame rate of video data and usually are set to small values such 10 pixels for S and 3 seconds for F.

For the case when two moving regions merge together, the corresponding layers for these regions will not be granted an allocation; instead a new layer is created for the new region. If objects merge only for a short time, such as when two separate subjects walk in different directions and pass by each other, the newly created layer during the merge will have a lifespan usually less then the threshold value F. Henceforth, this layer will be ignored. After the split, the tracking process continues normally and subjects are allocated to their original layers. This is because the centroid parameters of the original layers were updated during the merge. For the situation when objects split from one moving block, new layers are created for every new region, unless the object is divided due to partial occlusion by a thin structure such as a vehicle or walking person moving or walking behind a lamppost. Figure (4.3) shows the results of tracking multiple moving objects simultaneously. The video scene consists of two subjects walking separately, and one moving vehicle. The three moving objects are tracked successfully during their lifespan by the layering technique. For visualization purposes, moving regions of each layer extracted from the video sequences are combined together into one image such that each colour corresponds to a different frame.

4.2.3 Occlusion Handling

Occlusion when tracking moving objects occurs when part of the object or the object itself is hidden by another static or moving object and therefore becomes invisible in the scene. Tracking under occlusion is difficult as features like size and aspect ratio will be largely affected. Therefore the allocation cannot be performed. Firstly, the case of full occlusion where the object becomes totally invisible is handled by the centroid update function; every time the layer does not have a match, the algorithm updates the centroid parameters of the corresponding layer using the history data from the previous frames. Another aspect to consider is determining the duration of the time to wait for the object to reappear. A disappearance threshold D is defined such that if a layer does not obtain a match after D frames (typically set to 120 frames), then the layer is closed and will not accept further allocations. If the object reappears in the scene after the threshold duration has passed, then a new layer will be created for it. For the case of partial



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

FIGURE 4.3: Tracking Multiple Objects (a) Frame from the Video Sequence. (b) Layer 1: Moving Vehicle (c) Layer 2: Walking Person (d) Layer 3: Walking Person

occlusion, the matching allocation criteria described earlier is not robust for coping with occlusion as low-level features for the candidate objects are usually characterised by dramatic changes. The bounding box is used to handle partial occlusion including the case when moving regions are split by occlusion. If an unallocated candidate is mostly contained within the bounding box of a layer, then this object will be allocated to this layer as it belongs to it.

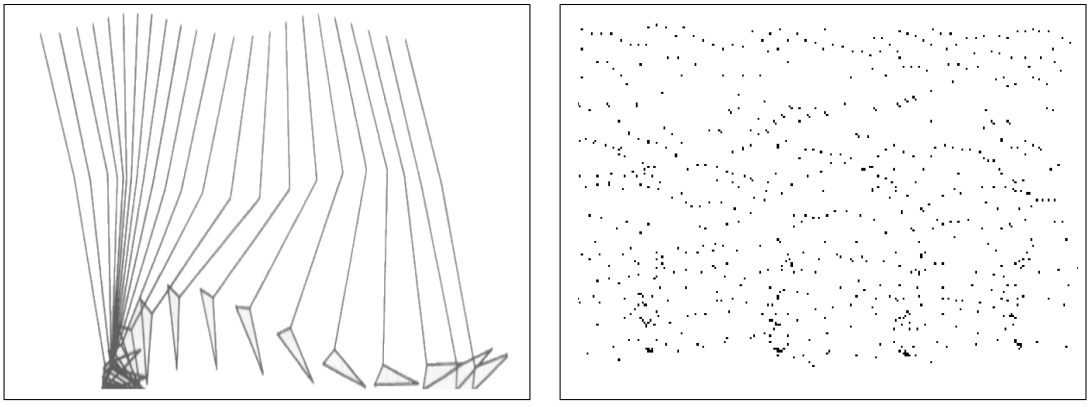
4.3 Object Classification

The proposed method classifies moving objects into either single walking subject, a group of people or undefined objects such as vehicles. The classification procedure is based on the analysis of the rhythmic pattern of gait motion. Because gait is a symmetric and

periodic motion, the distances between two close heel strikes, i.e., step lengths should be the same during all gait cycles. This cue is considered as the main feature to distinguish walking subjects from other moving *non-people* objects. Therefore, to detect walking pedestrians, the heel strikes are extracted for the purpose of classifying moving objects.

4.3.1 Heel Strike Extraction

During the striking phase, the foot of the striking leg stays at the same position for about half a gait cycle, whilst the rest of the human body moves forward as shown in Figure (4.4(a)). Therefore, if we apply a low-level feature extraction method (such as edge or corner operators) on every frame and then combine all the resulting frames together into a single frame, we will observe that dense regions are produced at the heel strike areas as depicted in Figure (4.4). Since the primary aim of this research is



(a) (b)

FIGURE 4.4: The Heel Strike Regions : (a) Position of the Right Leg during Walking [131], (b) Corner Image for Walking subject

the perception of human motion, we have chosen to use corners because they maintain sufficient information to analyse the human motion. Furthermore, a robust vision system based on corner detection can work for low-resolution applications. We have applied the Harris corner detector on every frame t. For every moving object belonging to the i^{th} layer, we take the corresponding corners for this object as the intersection of the object blob with the resulting frame. Afterwards, corner images corresponding to the same

object are combined together into a single image using the following expression (4.8):

$$C_i = \sum_{t=1}^{N} (H(I_t) \wedge L_{i,t}) \tag{4.8}$$

where H is the Harris corner detector, I_t is the original image at frame t, $L_{i,t}$ is i^{th} layer. The $\wedge$ operator is the logical conjunction function which returns either 1 or 0 for the logical values $true$ or $false$ consecutively. The expression $H(I_t) \wedge L_{i,t}$ is used to return only the corner points belonging to the i^{th} object at frame t. An example output of this process is illustrated in Figure (4.4(b)).

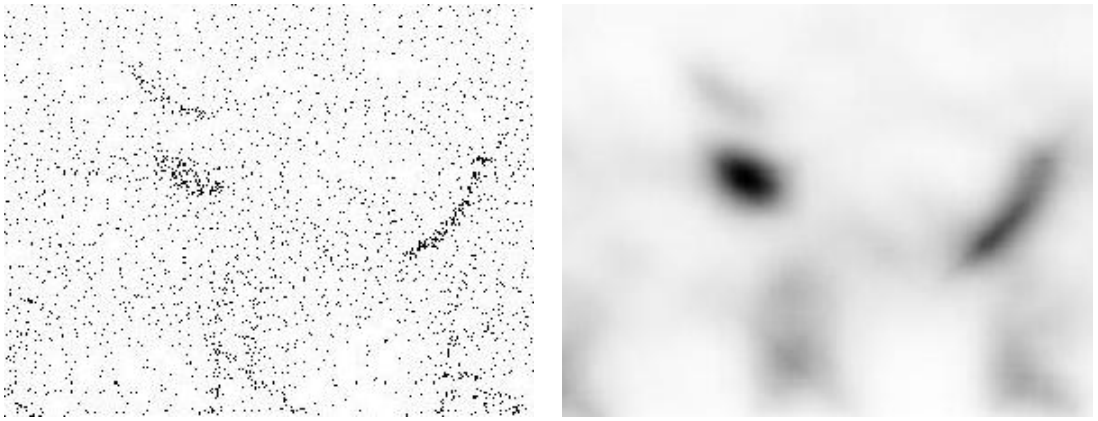
4.3.1.1 Computation of Proximity Image

Because the striking foot is stabilised for half a gait cycle, a dense area of corners is detected in the region where the foot strikes the ground. In order to locate these areas, we have devised a measure for point proximity in an image to find where the crowded region in a given image. The value of proximity at point p is dependent on the number of points within the neighbourhood region R_p and their corresponding distances from p. For simplicity, R_p is assumed to be a square area with the centre p, and width of $2r$ which can be determined as the ratio of the total image pixels to the total of corner points in C_i.

In order to compute the proximity image, we initially compute the neighbourhood proximity d_p for the region R_p corresponding to the point p, such that d_p is also a square region with the same width as R_p. The computation is carried out in an iterative process starting from the boundaries of R_p. It computes the nearness value of points with respect to the centre p and then it iterates inside and accumulate the previous computed values as expressed in the following equation:

$$\begin{cases} d_p^r = \frac{N_r}{r} \\ d_p^i = d_p^{i+1} + \frac{N_i}{i} \end{cases} \tag{4.9}$$

where d_p^i is the proximity value for rings of distance i away from the centre p, and N_i is the number of corners at distance i from the centre.



(a) (b)

FIGURE 4.5: Example Results for the Corner Proximity Measure: (a) Input Image, (b) Corner Proximity Image

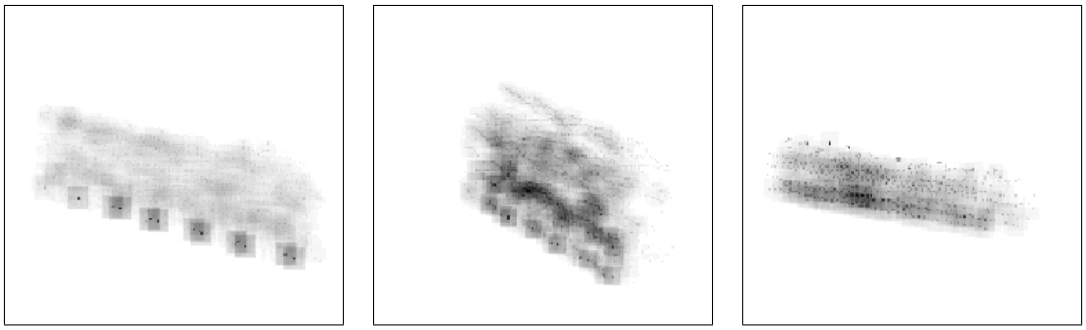
Afterwards, in order to produce the proximity image, we accumulate all the neighbourhood proximity values d_p for all points p into one image as described in the following equation:

$$D = \sum_{x=0}^{X} \sum_{y=0}^{Y} shift(d_{(x,y)}, x, y) \tag{4.10}$$

where X and Y are the width and height of the image respectively. $d_{(x,y)}$ is the neighbourhood proximity value for region $R_{(x,y)}$. The $shift$ function places the proximity value $d_{(x,y)}$ on a blank image of size $X \times Y$ at the position (x, y). An output of the point proximity for a sample image is shown in Figure (4.5). The input image contains a point cloud with a number of dense regions. The resulting image has darker areas which correspond to the crowded regions in the input image. Because it is a challenging task to formally determine which regions in a given image are crowded or dense as opposed to using a simple 2D histogram function, the problem becomes a question of detecting darker regions of the derived proximity image.

For the application context of this research, we have applied the point proximity measure on different moving objects being captured using a surveillance camera. Moving objects include a single walking individual, a group of two subjects walking together and

a vehicle. The results are presented in Figure (4.6). Clearly, the corner proximity image depicted in Figure (4.6(a)) for the walking subject originally shown in Figure (4.3), has a pattern of darker spots being detected at the bottom part of the image as the foot strikes the ground. Moreover, these darker regions are observed to have mostly the same level of darkness with consistent distance between two consecutive regions. On the other hand, the proximity image of people walking together constitutes a noisy pattern corresponding to the foot steps of subjects, however, the lower part of the proximity image is darker than the upper part of the image. For vehicles, the proximity image has almost a flat and consistent pattern with peaks located at random positions in the image.



(a) (b) (c)

FIGURE 4.6: The corners proximity images for : (a) Single Walking Person, (b) Group of People, (c) Moving Vehicle

4.3.1.2 Peak Detection

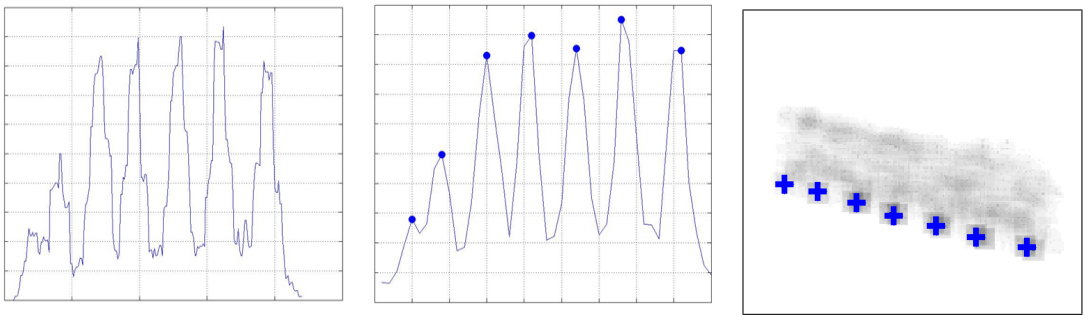
The extraction of the heel strike positions, i.e. the process of detecting darker regions in the proximity image can be applied using the K-means unsupervised clustering method. However, this technique requires the value of K to be known. Alternatively, we define the function f as the vertical projection of a point proximity image:

$$f(x) = \sum_{y=1}^{H} P(x, y) \tag{4.11}$$

where P is the proximity image and H is the image height. For noise reduction, the function f is smoothed using a low-pass filter. Afterwards, the local maxima points of

the smoothed graph are derived as they reflect the heel strike positions on the horizontal axis. This is performed by detecting the zero-crossings of the derivative function of f.

Having obtained the x coordinates of the heel strikes, the corresponding y coordinates can easily be approximated as the centroid of points along the vertical line which crosses the horizontal axis at the coordinate x. Figure (4.7) presents the results of heel strike extraction applied to the point proximity image shown in Figure(4.6(a)). The projection function f is plotted in Figure (4.7(a)). Figure (4.7(b)) shows the detection of local maxima from the smoothed function. The results of the extraction are shown in Figure (4.7(c)). Following this approach, the stride parameters of the walking people (i.e., the distance between two strikes of the same leg) and the gait step can be measured from the obtained results.



(a) (b) (c)

FIGURE 4.7: Extraction of Heel Strike Positions : (a) Vertical Projection of the Proximity Image. (b) Local Maxima Detection on the Smoothed function f. (c) Extraction Results

4.3.2 Derivation of Feature Vector

Clearly, the point proximity images for walking subjects show larger peaks at the bottom as legs have static periods. Furthermore, since gait is periodic, the stride length should be the same for different gait cycles, therefore the standard deviation of distances between two close strikes, i.e. peaks tend to zero. For the classification of moving objects, we define the feature vector $< \sigma, b, \alpha >$ where σ is the standard deviation value of distances between two successive peaks extracted from the point proximity image. The value of σ should tend to zero for walking subjects and becomes larger for moving vehicles. b is

the proportion of the lower part of the proximity image computed using the following equation (4.12):

$$b = \frac{\sum_{p \in B}}{\sum_{p \in C}} \tag{4.12}$$

where C is the proximity image and B is the lower half of the non-blank part of the proximity image C. The value of b should be larger for both single subject and a group of people as most peaks are located at the lower side of the proximity image. The feature α is the aspect ratio of height to width of the bounding box. This value is mainly used as a discriminative feature between a single subject and a group of people. To account for the scaling of the different dimensions, the data are normalised using:

$$f_n = \frac{f - \mu_f}{\sigma_f} \tag{4.13}$$

where such that f_n is the normalized feature. μ_f and σ_f are the mean and standard deviation of the feature f. The k-nearest neighbour rule is employed for the classification of moving objects based on the derived feature vectors.

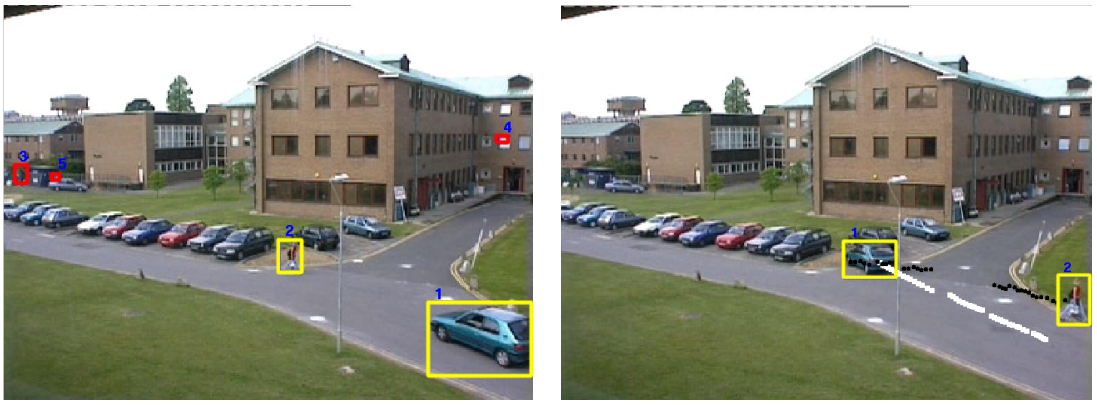
4.4 Results and Analysis

To demonstrate the efficacy of our approach for the use of automated visual surveillance, the proposed methods for tracking moving objects and detecting walking pedestrians have been evaluated on publicly available data which contains a variety of scenarios and conditions. The evaluation experiments are applied on a set of four video sequences provided by PETS 2001[1], compressed in JPEG format. The videos are recorded at real-time frame rate (25 frames/second) and they are filmed in an unconstrained outdoor environment with walking people and moving vehicles. The size of the video frames is reduced to 384x288.

[1] Available from the University of Reading at : http://ftp.cs.rdg.ac.uk/PETS2001/

4.4.1 Tracking and Occlusion Analysis

The presented algorithm for tracking multiple moving objects is tested on the set of video sequences. Moving objects are tracked successfully during their life span in the monitored scene. However, it is observed that the foreground segmentation algorithm is detecting many false regions due to background clutter (as moving tree leaves), JPEG artefacts and other environmental conditions. This problem is handled efficiently by the use of the size and lifespan threshold filters as most of the false regions have shorter lifespans or smaller sizes, henceforth they are ignored by the tracking algorithm. Figure (4.8) presents an example of tracking multiple moving objects. At the initial frame shown in (4.8(a)), there are two moving objects in the scene shown inside yellow bounding boxes, but the foreground segmentation algorithm has detected five different moving blobs of which three are false regions. The algorithm tracks successfully the moving regions as shown in Figure (4.8(b)) and ignores false regions due to either their short lifespan or small size. False candidate moving objects are shown inside red bounding boxes.



(a) (b)

FIGURE 4.8: Results of Tracking Multiple Moving Objects: (a) Initial Frame. (b) Tracking Results after 150 Frames

Furthermore, the tracking algorithm is capable of handling occlusion efficiently, and reallocating the occluded object to the correct layer when occlusion vanishes. Figure (4.9) shows a walking person partially occluded by a street light pole. The subject is detected as multiple separate moving regions by the foreground segmentation process as shown in Figure (4.9). The algorithm successfully allocates the detected blobs to the

layer corresponding to the subject because they are not allocated to existing layers and are mostly contained within the predicted bounding box of the walking subject. After occlusion, tracking is carried out successfully as shown in Figure (4.9(f)).

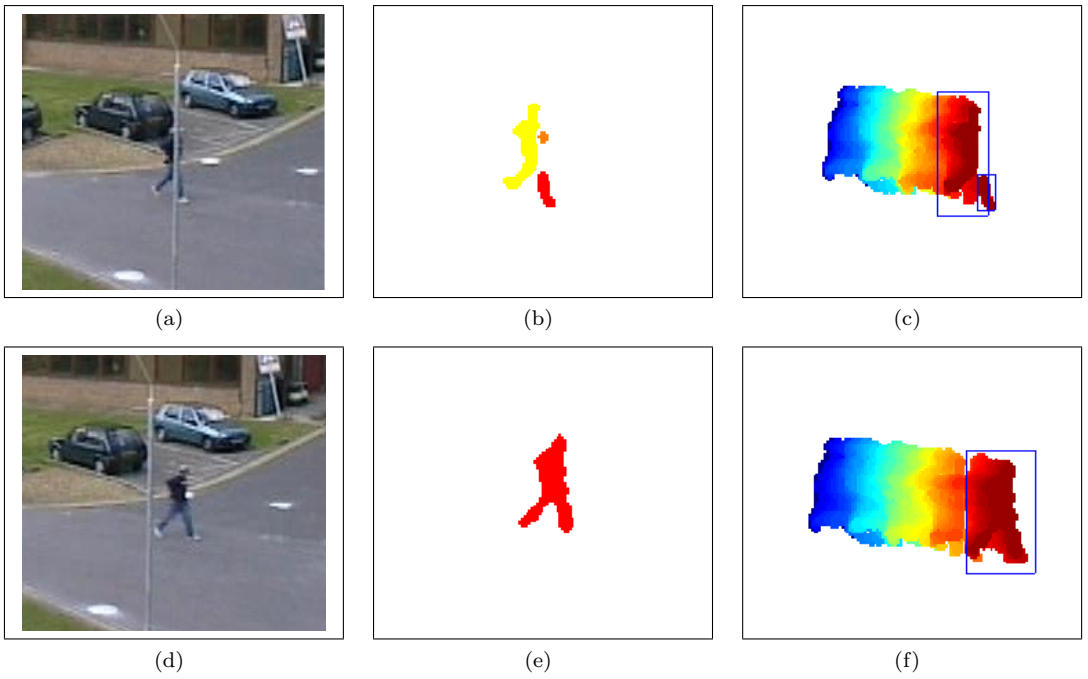


FIGURE 4.9: Experimental Results for Handling Occlusion : (a) Walking subject being occluded. (b) Foreground segmentation. (c) Allocation of moving regions into their layers.(d) Walking subject after occlusion. (e) Foreground segmentation of d. (f) Tracking recovery of the walking subject after occlusion.

4.4.2 Classification Results

To verify the effectiveness of our approach to classify moving objects using their gait pattern, we have carried out a number of experiments on the PETS video data containing a total of 27 moving objects from the set of four video sequences. The leave-one-out validation rule is used to assess the performance of the classification using the k-nearest neighbour classifier. The system is able to discriminate between a single walking subject, a group of people and vehicles efficiently using the proposed features, achieving a Correct Detection Rate of %100. The results of the classification are detailed in Table (4.1). The feature vectors of moving objects are projected into the feature space shown in Figure (4.10). It is clearly revealed that the standard deviation feature of the stride parameters,

is a strong cue to distinguish between walking people and vehicles. This is consistent with the findings of BenAbdelkader *et al.* [12] where the stride parameters are utilised for people identification.

TABLE 4.1: Moving Objects Classification Results

Type of Object	Number of instances	Instances Correctly Classified
Single Person	15	15
Group of People	5	5
Moving Vehicles	7	7

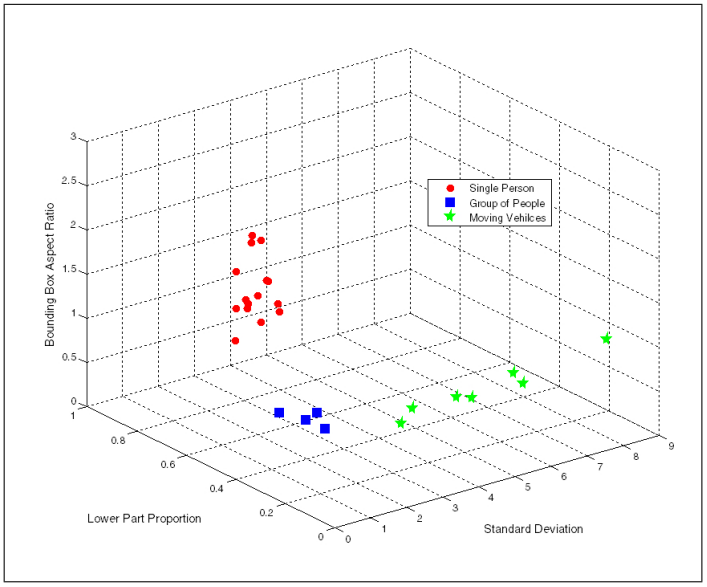
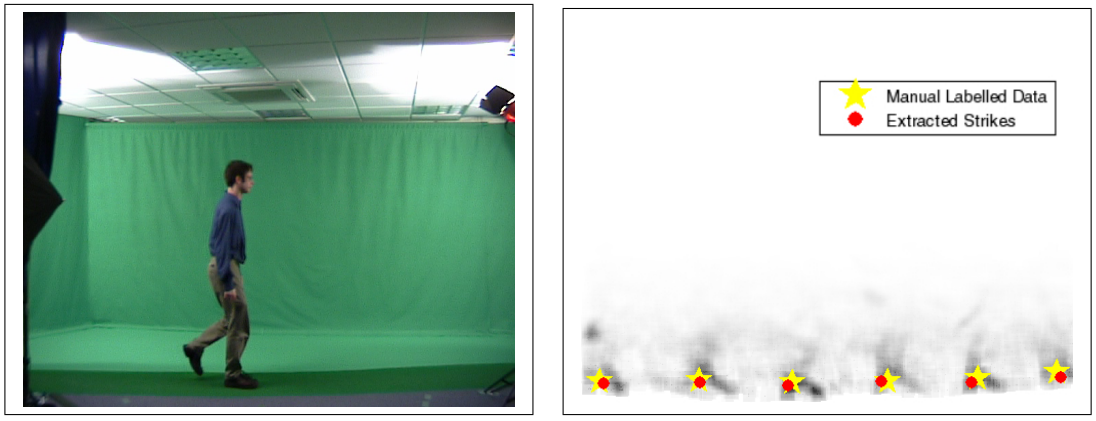


FIGURE 4.10: Feature Space for Moving Object Classification

4.4.3 Performance Analysis for Heel Strike Detection

Although the classification results of moving objects are promising, further experiments are conducted to confirm the robustness of the proposed method for extracting the heel strikes. The algorithm is applied on a set of 100 different subjects from the SOTON database. All subjects are filmed in an indoor environment with controlled conditions. A total of 510 strikes are extracted successfully from a total of 514 strikes with only four strikes being missed by the algorithm. The mean error for the positions of the strikes extracted by the algorithm compared to strikes manually labelled is %0.52 of the person's height. The error is measured using the only available and consistent error

metric by normalising the Euclidean distance between the manual and the extracted data to subject's height. The height is measured by taking the average height of the silhouettes across the video sequences. Figure (4.11) shows the results of heel strike extraction by the described method compared with the data labelled manually for one video sequence and it can be observed that the match is indeed close.

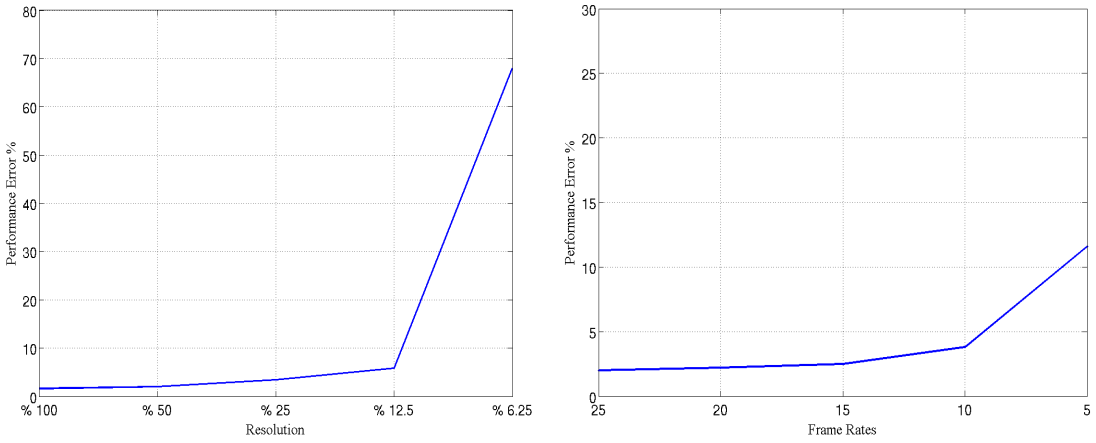


(a) (b)

FIGURE 4.11: Experimental Results for Heel Strikes Extraction: (a) Walking subject. (b) Extracted strikes compared with data manually labelled

The performance of the heel strike detection algorithm is investigated on different situations including low-resolution images as well as video sequences with low frame rates. Figure (4.12(a)) presents the performance error of the algorithm for deriving the heel strikes from varying image resolutions. The error is measured as the sum of false detected and missed strikes. The size of the video sequences are reduced gradually from an original size of 576x720 pixels such that the aspect ratio is kept constant. The graph shows that the algorithm is still able to derive the heel strikes at a resolution of 72x90 with an acceptable error rate. However, the algorithm fails when the resolution is reduced to 36x45 pixels where the algorithm misses the extraction of most strikes. Figure (4.13) shows examples of the heel strike detection algorithm at various resolutions. Furthermore, the performance error is also evaluated by decreasing the frame rates of the video stream resized at resolution of 288x360 pixels as depicted in Figure (4.12(b)). This is performed by dropping a number of frames. It is observed that the algorithm

performance is not much affected even when dropping 60% of the frames (i.e., frame rate at 10 frames/second).

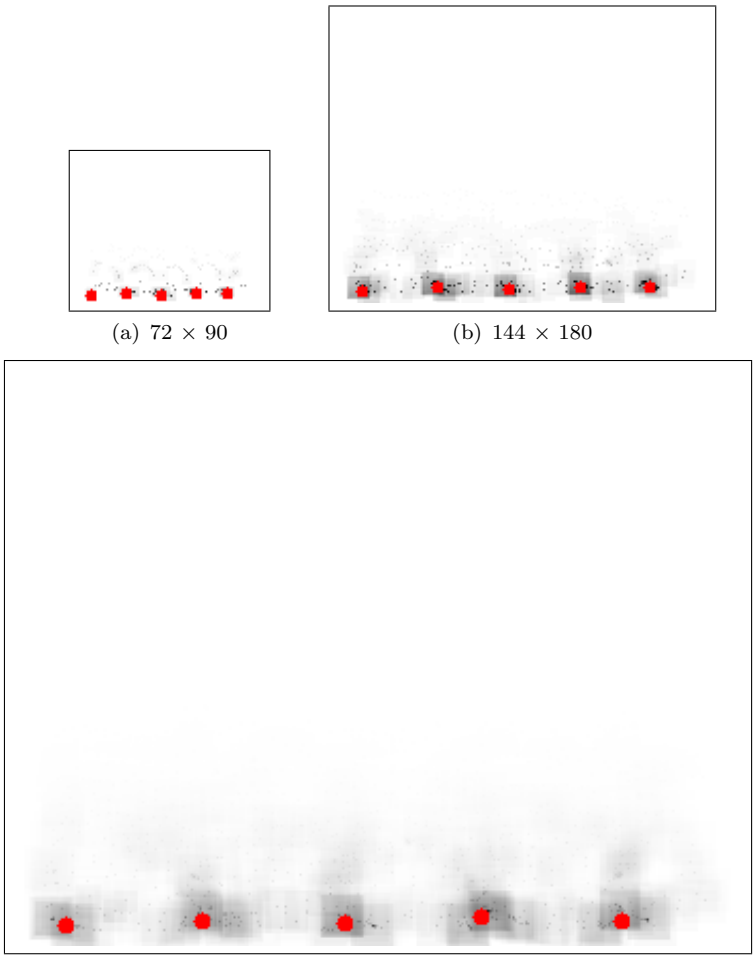


(a) (b)

FIGURE 4.12: Performance Analysis for Heel Strikes Detection.

4.5 Conclusions

We have proposed a new method to classify moving objects for automated visual surveillance. Multiple objects are tracked successfully through the use of shape-based parameters to allocate them to different layers. Problems encountered during tracking such as background clutter, appearance of uninteresting objects and entry and exit of objects are handled efficiently. Finally, moving regions are classified into either a single walking person, a group of people or an undefined object such as a vehicle. In contrast to approaches that employ shape-based parameters for classification, we have explored an alternative technique for walking people detection based their gait motion.



(a) 72 × 90

(b) 144 × 180

(c) 288 × 360

FIGURE 4.13: Heel Strike Detection at Various Resolutions.

Chapter 5

Feature Extraction via Model-Based Method

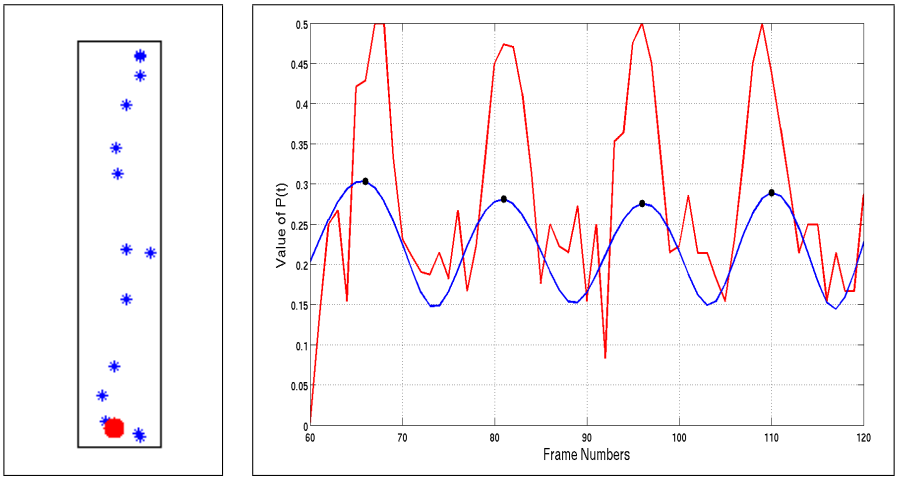
Psychological experiments confirmed that joint motions embed sufficient information for the perception of human behaviour, gender classification as well as identity recognition. Nevertheless, the extraction of the joint trajectories has proved challenging using marker-less methods. This chapter describes the proposed model-based method deployed to derive the positions of the joints of walking subjects. Motion models describing the gait motion are produced via gait analysis performed on data collected from manual labelling. We present the recursive evidence gathering algorithm which is utilised for the localisation of the joints. Elliptic Fourier Descriptors are used to represent the motion models in a parametric form. Experimental results for people recognition using gait are drawn at the end of the chapter.

5.1 Gait Periodicity Estimation

Periodicity is one of the significant characteristics of the gait pattern. The computation of gait period and cycle partitioning are crucially important steps for gait recognition algorithms. Several methods have been proposed to estimate gait periodicity as it provides essential information for the extraction of gait features. Typically, gait cycles are detected using either the periodic variation of the width for the silhouette bounding box or auto-correlation methods. Cutler *et al.* [28] proposed a real-time method for measuring gait periodicity based on self-similarity. BenAbdelkader [13] used the variation of the bounding box width for the subject silhouette to measure the gait periodicity. In [113], Chellappa *et al.* proposed an adaptive filter used to filter the foreground sum signal prior to the measurement of the gait cycles using the minima of the signal. However, most of the gait cycle detection algorithms suffer from accuracy problems due to the varying walking speed as well as badly segmented silhouettes.

In order to avoid the foreground segmentation problems, the estimation of gait periodicity is based on the derived heel strike data. The gait cycles are partitioned at the start of the striking and terminal stance phases for each leg. Henceforth, the gait periodicity can be estimated as the absolute difference between the frame numbers of two consecutive heel strikes of the same leg. Based on gait analysis, it is observed that during the start of the terminal stance phase, most of the human body is vertically projected into the striking region as shown in Figure (5.1(a)). We introduce the rectangular-based region R_i such that its width is set to the side-view width of subject experimentally estimated as a quarter of the distance between two consecutive strikes $\{i^{th}, (i+1)^{th}\}$, and it is positioned at the striking point i^{th}. The height of R_i is defined as the height of the walking subject.

Based on the observation of gait motion discussed in this section, a histogram-based function P is derived for the partitioning of gait cycles. The value of P at a given frame, is defined as maximum values of the proportions of corners detected inside the different regions $R_{i,i+1,...}$ to the number of corner points detected within the frame. The function



(a) (b)

FIGURE 5.1: Gait Periodicity Detection : (a) The Start of the Terminal Stance Phase, (b) The smoothing of the Histogram function $P(t)$.

P is expressed in the following equation:

$$P(t) = \arg \max_{i \in strikes} \left(\frac{\sum_{p \in R_i} H(t,p)}{\sum_{p \in I} H(t,p)} \right) \tag{5.1}$$

where the function $H(t,p)$ returns 1 if the point p at the t^{th} frame is a corner and returns 0 otherwise. I is the the intensity image. Since we have used corners as the low-level feature operator, it happens that at the start of the terminal stance phase, the number of corner points of the segmented walking subject detected inside the regions R_i will become larger as the whole human body is almost contained within one of the regions R_i. Henceforth, the local maxima of the function P should reflect the frame numbers (timing) of the corresponding terminal stance phases. A low-pass filter is applied to the function P for noise reduction. The zero crossing is applied to the derivative of P to extract the local maxima and minima. Figure (5.1(b)) shows the smoothed function P. Indeed, human gait is a periodic motion as the function P has the form of a sinusoid function. The minimum points on the graph indicate the start of the striking phases.

5.2 Shape Extraction via Evidence Gathering

5.2.1 The Hough Transform

The Hough transform [54] was originally proposed to detect straight lines in an image using the slope-intercept parametrisation of a line. The transform was extended by Duda *et al* [38] to use a bounded parametrisation of a line instead, as the slope-intercept parameters have an infinite range causing the algorithm to be impractical for most cases. The main concept of the algorithm is that at a given point in the image, there is a finite number of potential lines that pass through the point. Each of these lines which is uniquely defined by its two polar parameters will be given a vote. The aim of the transform is to determine which of these potential lines pass through most points in the image, i.e., the line which has the greatest number of votes accumulated.

Furthermore, Duda and Hart [38] described how the Hough Transform can be used to detect any analytically defined shape, giving an example of the algorithm being used to detect circles. The essence of the Hough transform is to determine the free parameters defined in a multidimensional space where each dimension represents a parameter [6]. Feature points are matched to the parametric analytic representation of the shape to find the mapping between the locus (i.e., feature point or point of interest.) and the free parameters and hence, gather evidence by adding votes to the multidimensional accumulator space for the corresponding parameters. The Hough Transform is proved an efficient template matching algorithm providing optimal results even in noisy conditions or where there are gaps in the boundary of the shape [6, 48].

Ballard [9] proposed the General Hough Transform (GHT) to detect arbitrary shapes with scale and rotation invariance. The GHT replaces the analytic parametric constraints with a non-analytic tabular representation. The table describes the position of feature points in the shape relative to a reference point and is indexed by the gradient information at each feature point. However, the GHT suffers from aliasing and rounding errors when scaling and rotating the shape [5]. Alternatively, the Elliptic Fourier

Descriptor was then used to mitigate the drawbacks of the GHT by giving a continuous representation of the original shape [6].

5.2.2 Shape Parametrization using Fourier Descriptors

Fourier theory has been used for the analysis of curves and boundaries of shapes for several years. Fourier analysis provides a means for extracting features or descriptors from images which are useful characteristics for image understanding. These descriptors are defined by expanding the parametric representation of a curve in Fourier series providing a compact and accurate representation of the curve. The main motivations for using the Fourier Descriptors are the flexibility, simplicity in computation as well as the vectorial parametrization [6] which is explained further in this section.

Let f be the function for the boundary of a given shape; the function f can be represented using the Elliptic Fourier Descriptors [6, 47], where the Fourier series is based on a curve expressed by a complex parametric form as follows:

$$f(t) = x(t) + iy(t) \tag{5.2}$$

where $t \in [0, 2\pi]$. $x(t)$ and $y(t)$ are approximated via the Fourier summation by n terms:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \tag{5.3}$$

such that a_0 and b_0 define the position of the shape's centre, and $X(t)$ and $Y(t)$ are computed as:

$$\begin{aligned} X(t) &= \sum_{k=1}^{n} a_{x_k} cos(kt) + b_{x_k} sin(kt) \\ Y(t) &= \sum_{k=1}^{n} a_{y_k} cos(kt) + b_{y_k} sin(kt) \end{aligned} \tag{5.4}$$

where a_{x_k},a_{y_k}, b_{x_k} and b_{y_k} are the set of the elliptic coefficients which can be computed by a Riemann summation [6] shown as follows:.

$$\begin{cases} a_{xk} = \frac{2}{m} \sum_{s=1}^{m} x_s \cos(k_s \frac{2\pi}{m}) \\ a_{yk} = \frac{2}{m} \sum_{s=1}^{m} y_s \cos(k_s \frac{2\pi}{m}) \\ b_{xk} = \frac{2}{m} \sum_{s=1}^{m} x_s \sin(k_s \frac{2\pi}{m}) \\ b_{yk} = \frac{2}{m} \sum_{s=1}^{m} y_s \sin(k_s \frac{2\pi}{m}) \end{cases} \tag{5.5}$$

such that m is the number of points (x_s, y_s) in the shape model shape. The value of k represents the number of ellipses that formulate the shape. According to the Nyquist sampling theorm, k ranges between 1 to $m/2$. The larger values of k, the more accurate representation is reconstructed for the shape model.

For a representation invariant to rotation and scaling, we need to represent f in a parametrized form to cover all the possible graphs or shapes which can be derived by applying appearance transformation to the function f including rotation and scaling. Henceforth, the x and y components of function f can be rewritten in the following parametric form:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} cos(\alpha) & -sin(\alpha) \\ sin(\alpha) & cos(\alpha) \end{bmatrix} \begin{bmatrix} X(t) * s_x \\ Y(t) * s_y \end{bmatrix} \tag{5.6}$$

where α is the rotation angle, s_x and s_y are the scaling factors across the horizontal and vertical axes respectively. Equation (5.6) leads to a concise complex notation for f:

$$\begin{cases} f = T + R_\alpha \left(s_x X(t) + s_y Y(t) i\right) \\ T = a_0 + b_0 i \\ R_\alpha = \cos(\alpha) + \sin(\alpha) i \end{cases} \tag{5.7}$$

Based on the final parametric form of f shown in Equation (5.7), given the values of X and Y of an enclosed shape or contours defined as F that are pre-computed using equation (5.4), the derivative shapes of F that are produced via scaling, translation

as well as rotation, can be represented using six parameters: a_0, b_0, α, s_x, s_y and $t \in [0, 2\pi]$. In fact, the number of free parameters needed for the Hough Transform is totally independent of the complexity of the shape which is defined using the elliptic Fourier Descriptors. This is because the defined parameters are related to the appearance transformations which define all the shapes that can be derived form the original shape [6]. Figure (5.2) shows a shape represented using the parametrised form described in equation (5.7) where we have applied a number of transformations including rotation and scaling.

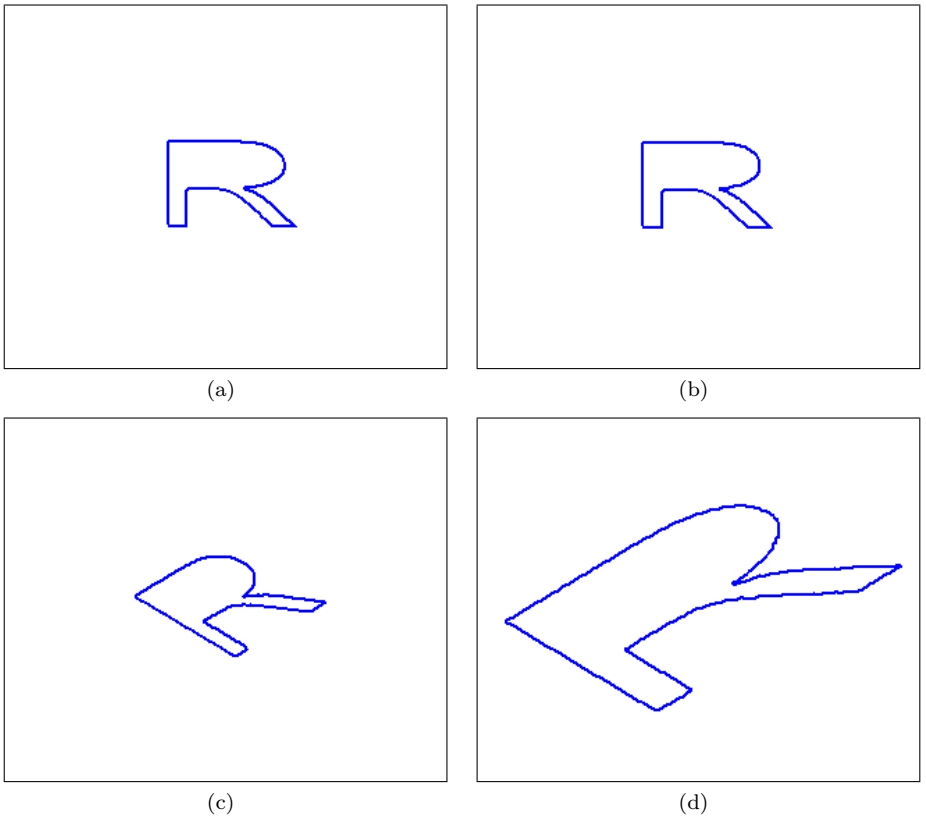


FIGURE 5.2: Parametric Representation of shapes using Elliptic Fourier Descriptors (EFD): (a) Original Shape. (b) Shape produced using the EFD. (c) Rotated Shape. (d) Rotated and Scaled Shape

5.2.3 Recursive Evidence Gathering Algorithm

The recursive evidence gathering algorithm is proposed for the purpose of obtaining accurate localisation of the joint trajectories. The evidence gathering process used for the

extraction of articulated objects is applied in conjunction with the Hough Transform consisting of two phases: i) global pattern extraction, and ii) local feature extraction. The former phase is aimed to find the overall motion pattern based on a predefined model. In this case, the motion pattern is the spatial path of the joint positions over a sequence of frames. In this study, Elliptic Fourier Descriptors are used for the parametrisation of motion models. The Hough Transform is employed to determine the free parameters through the matching process of feature points across the whole sequence to the parametric function, and increase votes in the accumulator space accordingly. The parameters are then determined as the index or key of the accumulator space with the largest value. In the latter phase of the evidence gathering process, an exhaustive local search is performed within every frame to locate the features (i.e., joint positions) whereby, the local search is guided by the motion pattern extracted during the first stage to limit the search area.

Algorithm 5.2.1: RECURSIVEEVIDENCEGATHERING$(model, points)$

comment: Global Pattern Extraction

$Acc \leftarrow Array()$

for each $p \in points$

do $\begin{cases} parameters \leftarrow HoughTransform(model, p) \\ Acc[parameters] \leftarrow Acc[parameters] + 1 \end{cases}$

$Best \leftarrow indexOfMax(Acc)$

comment: Local Feature Extraction

$Traj \leftarrow Array()$

for each $r \in Frames$

do $\{Traj[r] \leftarrow SearchLocally(r, Best)$

comment: Recursion

if $Traj \equiv points$

then return $(Traj)$

else $\begin{cases} Traj = RecursiveEvidenceGathering(model, Traj) \\ \textbf{return } (Traj) \end{cases}$

For the recursive evidence gathering process, the algorithm initially derives the global motion pattern using all the possible points. Afterwards, local feature extraction is applied to every frame to locate possible candidate feature points (i.e., joint positions). In the following iterations, the algorithm uses only the possible candidate points detected in the previous iteration as the feature points for the extraction of a global pattern. The recursion will continue until the extracted motion pattern reaches a constant state. A pseudo-code for the recursive evidence gathering algorithm is detailed below with its two different stages.

5.3 Extracting the Joints' Positions

As discussed in Section (3.1.2.2), feature extraction is the core part of a marker-less motion capture system. This is mainly because the crucial information required for the perception of human motion are derived during this stage. Most of the model-based methods proposed for extracting gait features rely on angular motion models from medical studies describing the knee and hip angular movements. These methods derive the angular measurements of walking subjects without extracting the joint positions [122, 138, 27]. However, total reliance on angular models for feature extraction has a number of drawbacks including clothing type and carrying conditions which certainly affect the extraction accuracy. In this research, we focused on the extraction of the joints (knee, hip and ankle) using spatial motion models derived from gait analysis describing the displacement movement of the joints in the XY plane. The recursive evidence gathering algorithm is employed to locate the joint positions of walking subject.

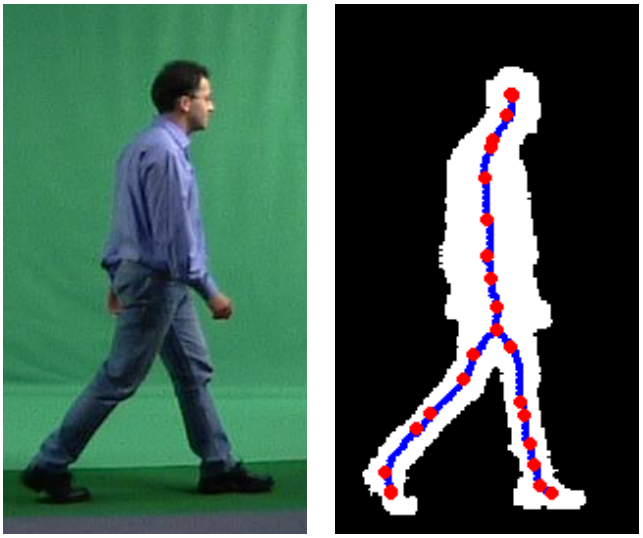
For the extraction of the global motion patterns for the hip, knee and ankle joints, spatial models are derived as the mean pattern from manual analysis of gait described in Section 3.3.5. Because only closed and continuous contours can be represented by the Elliptic Fourier Descriptors [71], we have converted the model templates describing the joints spatial motions into a closed cyclic contour by looping back along the curves. Let F be the model template. The Hough Transform [54] is used to search for the best parameters to obtain a shape that is similar to the model graph F and fits through most

of the feature points in the video sequence. However, to use the Hough transform with the spatial models represented via the parametric form described in Equation (5.7), a five-dimensional space is required. Thus, the algorithm would be computationally intensive and infeasible to implement for such application. In spite of the fact that some methods were proposed to reduce the processing time of the Hough Transform [75, 6], the computational load of these methods does not meet the requirements of most applications [6]. Alternatively, the heel strike data could be incorporated to reduce the complexity of the parameter space and therefore, dramatically reduce the computational cost.

In order to search for the motion patterns of the different joints using the evidence gathering approach, corner points are derived as the basic low-level features which are matched against spatial templates to deduce the best parameters. To extract candidate joint positions, a skeletonization process is performed on the silhouettes of walking subjects. The Distance Transform [88] is applied to derive the skeleton from the silhouette images. Afterwards, the Harris corner detector is applied to extract the feature points as the points of high curvature on the skeleton of the subject. The feature points include also junctions as well as the end points of the skeleton. Figure (5.3) shows the extraction of the feature points from silhouette data using the skeletonization process and the Harris corner detector. A number of corners points are detected along the skeleton because of the changes of curvature. In order to improve the localisation of the motion patterns as well as reduce the computational cost, anatomical knowledge of the body segment properties [32, 97] are used to filter out the irrelevant candidate joints. For instance, the points detected in the upper section of the silhouette are ignored as the hip, knee and ankle joints are located in the lower section.

5.3.1 The Ankle Joint

The model template describing the spatial motion of the ankle joint is produced from gait analysis of manually labelled data. The model template is taken as the mean of the labelled data for one gait cycle starting from the striking phase. Figure (5.4) depicts



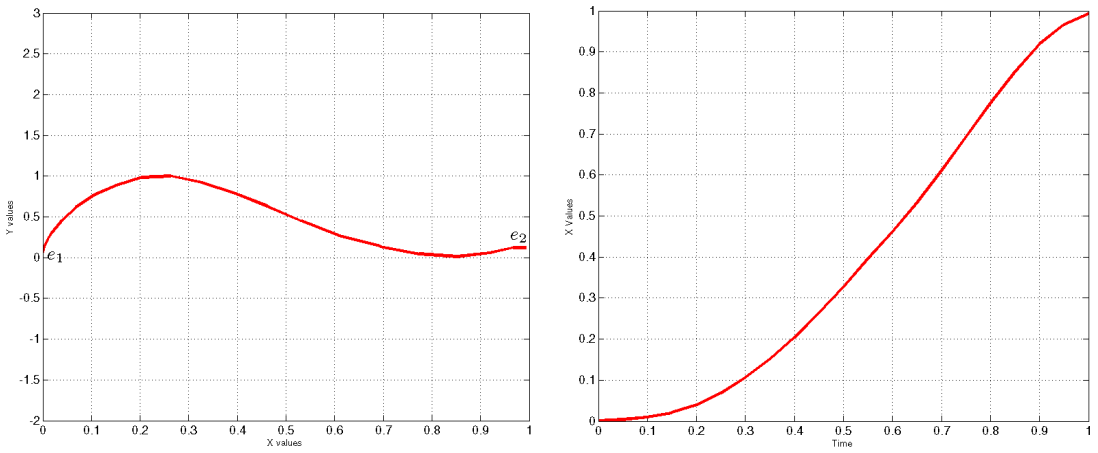
(a) (b)

FIGURE 5.3: Extraction of Candidate Points using the Skeletonization Process: (a) Walking Subject. (b) Extraction of Joint Candidate Points.

the spatial model and the horizontal displacement of the ankle. In order to extract the ankle pattern across consecutive frames using the evidence gathering algorithm, the five-dimensional space required for the parametrisation of the model template defined in Equation (5.7), is reduced to only one parameter s_y as expressed in Equation (5.8). Because the heel strike points are known and lie on the ankle motion pattern as its two end points, this knowledge is used to deduce the two parameters s_x and α as the distance and angle between the two strikes respectively. The parameter reduction procedure is explained in Appendix A.

$$s_y = \frac{(y - y_{s_n}) - (X(t) - x_e)s_x \sin(\alpha)}{(Y(t) - y_e)\cos(\alpha)} \tag{5.8}$$

(x, y), (x_e, y_e) and (x_{s_n}, y_{s_n}) are the coordinates of the locus point, left end point of the model template and the n^{th} strike respectively. The Hough Transform is applied to search for the curve which resembles the ankle spatial template. The transform iterates through the locus points detected by the Harris corner detector to compute the s_y parameter (the scale through the vertical axis) and to increment the voting accumulator array for the corresponding index. The translation parameters are then computed after



(a) (b)

FIGURE 5.4: Template model and horizontal displacement of the ankle motion: (a) Template model for the ankle motion. (b) Horizontal displacement of the ankle.

determining the best value of s_y using the following equation:

$$T = s_n - (\cos(\alpha) + i\sin(\alpha))(x_e s_x + i * y_e s_y) \tag{5.9}$$

Figure (5.5) shows an output of the global pattern extraction of the ankle joints for both the right and left legs for a walking subject.

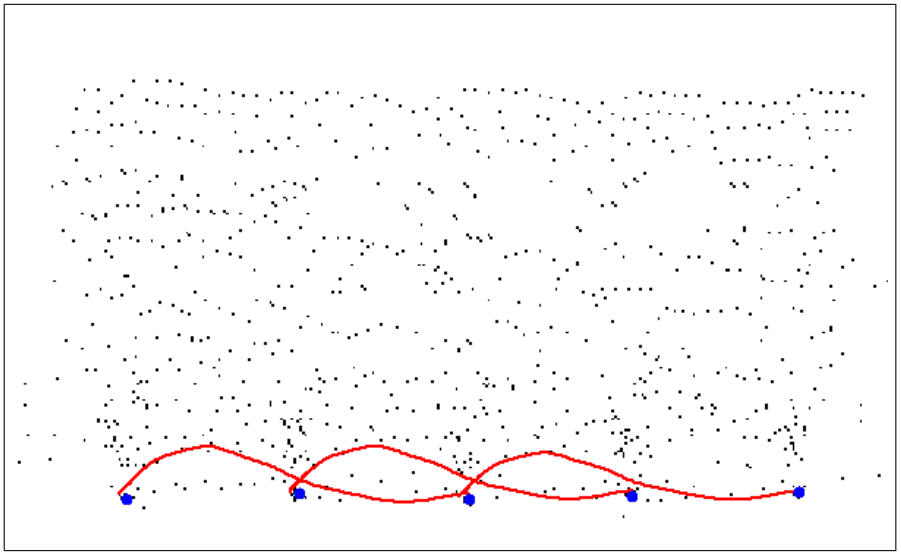


FIGURE 5.5: Global Pattern Extraction of the Ankle Joint.

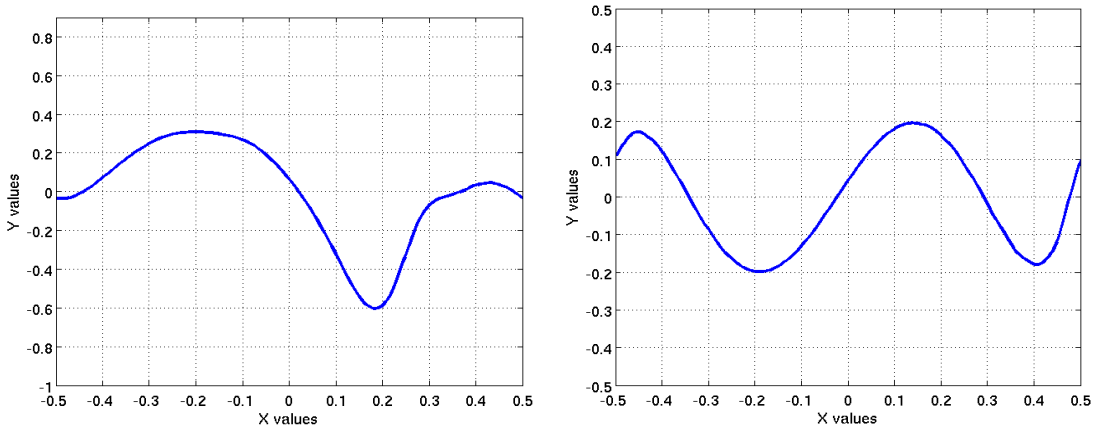
In the second stage of the evidence gathering algorithm, local search is performed

within every frame to derive the precise position of the ankle. People more or less have the same velocity pattern for the ankle horizontal displacement when walking as concluded in Section 3.3.5. Therefore, a displacement model is produced and used to estimate the distribution of the x values through the temporal data. Let N be the number of frames between the two strikes of the same foot s_n and s_{n+2}. The x coordinates of the ankle trajectories can be approximated as expressed by the following equation:

$$x_f = s_x V\left(\frac{f}{N}\right) + x_{si} \tag{5.10}$$

where f is the frame number, s_x is the horizontal scale parameter. x_{si} is the x coordinate of the i^{th} heel strike. V is the horizontal displacement model function of the ankle joint derived from manually labelled data. Afterwards, the y coordinates of the joints are approximated using the global motion pattern derived during the first stage. For accurate localisation of the ankle joint at a given frame f, an exhaustive local search is conducted in the area whose centre is the approximated location i.e. (x_f, y_f) in order to locate the ankle position as the nearest point to the motion pattern.

5.3.2 The Knee and Hip Joint



(a) (b)

FIGURE 5.6: Spatial Motion Templates: (a) Knee Joint. (b) Hip Joint.

For the extraction of the knee and hip trajectories, we apply the same methodology used for the localisation of the ankle. The mean templates for the spatial motion and

displacement of the knee and hip are derived for one gait cycle as shown in Figure (5.6). The gait cycle is set to start from the terminal stance phase when the leg is vertically straight such that the x coordinates of the knee and hip are almost the same as the heel strike. In order to reduce the dimensionality of the parametric model, the x coordinates of the heel strikes are exploited to deduce the horizontal scale and rotation angle parameters. The global search for the joint pattern is represented using only two parameters which are s_y and y_{s_n} as expressed in the following equation:

$$s_y = \frac{(y - y_{s_n}) - (X(t) - x_e)s_x \sin(\alpha)}{(Y(t) - y_e)\cos(\alpha)} \tag{5.11}$$

where y_{s_n} is the y coordinate of the point whose x coordinate is x_{s_n}. The procedure for parameter reduction is described in Appendix (A). An output of the evidence gathering algorithm showing the global extraction of the motion patterns for the knee and hip joints is presented in Figure (5.7).

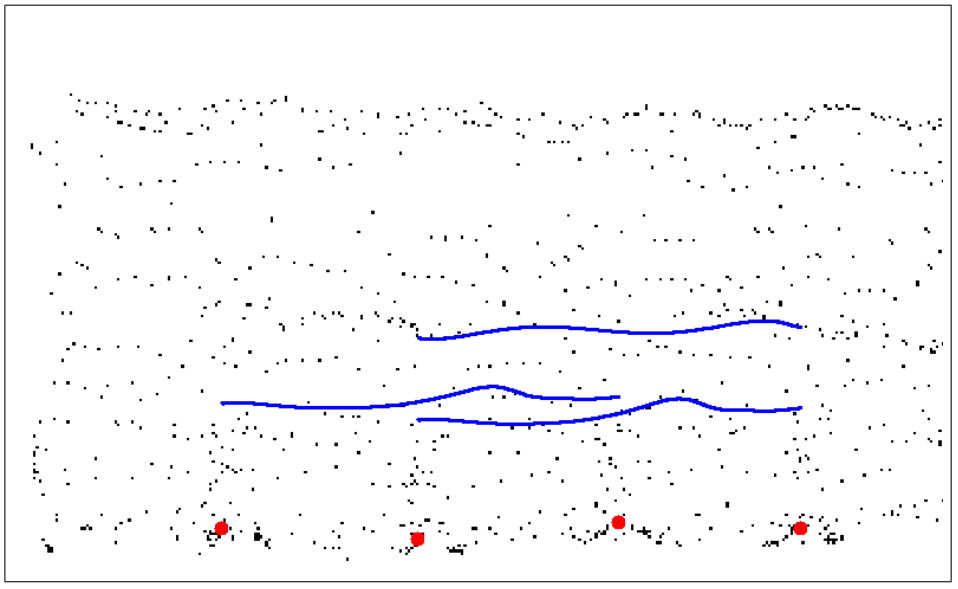


FIGURE 5.7: Global Pattern Extraction of the Knee and Hip Joint.

In order to obtain accurate results for the localisation of the knee and hip, additional constraints are applied to filter false locus points. This is due to the hard nature of the problem which is also faced even during manual extraction of the two joints. Besides the use of displacement motion model to locally search for the joints, the knee angle and the

mean x coordinate of the trunk are extracted from the skeletons of walking subjects and employed to enforce the accuracy of the extraction process for the knee and hip joints respectively. The x coordinate of the trunk is derived based on the use of anatomical knowledge whereby only the points located at the upper part of the human body are considered. The knee angle is approximated using linear regression of the skeleton points detected in the region above the extracted ankle joint as follows:

$$\theta_{knee} = \tan^{-1}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}\right) \tag{5.12}$$

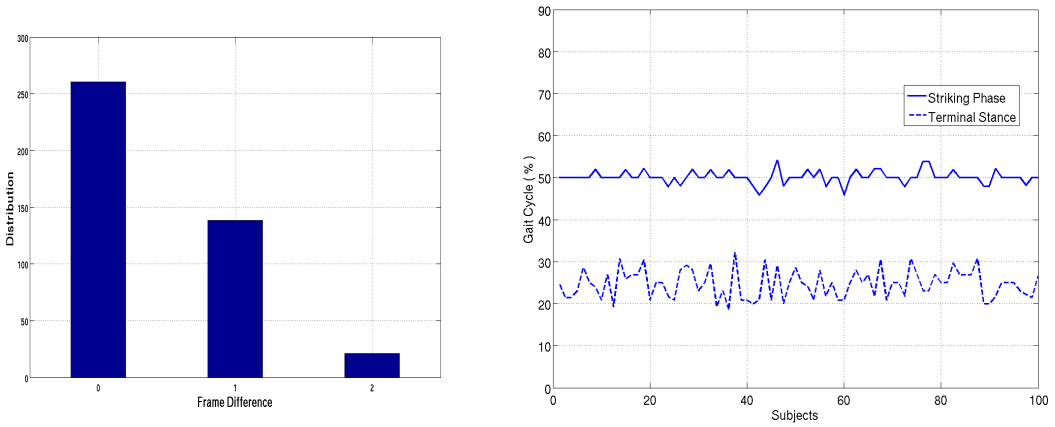
Where $(\bar{x}, \bar{y})$ are the centroid point for the (x, y) points. n is the number of points.

5.4 Experimental Results and Analysis

5.4.1 Periodicity Detection

In order to evaluate the accuracy of the gait cycle partitioning approach proposed in Section 5.1, we used 100 video sequences of walking subjects taken from the SOTON gait database. The frame numbers corresponding to the start of the striking phases which are extracted by the partitioning method are compared against ground truth data that have been manually collected and stored in XML format. The extraction error is computed as the average error of the absolute differences between the ground truth data and the extracted frame numbers of the striking phases. The error is approximated to ± 0.4 frames based on 420 comparisons. Figure 5.8(a) depicts the comparison results of automatic and manual detection of gait periods as a distribution chart of the absolute difference between automated and manual extraction. The gait period is estimated as the difference between the frame numbers of two successive strikes of the same leg.

To assess the consistency of the results derived by the gait partitioning algorithm with medical studies, the different gait phases including the terminal stance and the striking phase which are detected by the partitioning algorithm are derived for one gait cycle starting from the striking phase of the right leg. The results of the gait partitioning



(a) (b)

FIGURE 5.8: Gait Periodicity Detection: (a) Automated vs Manual Extraction. (b) Detection of Gait Phases

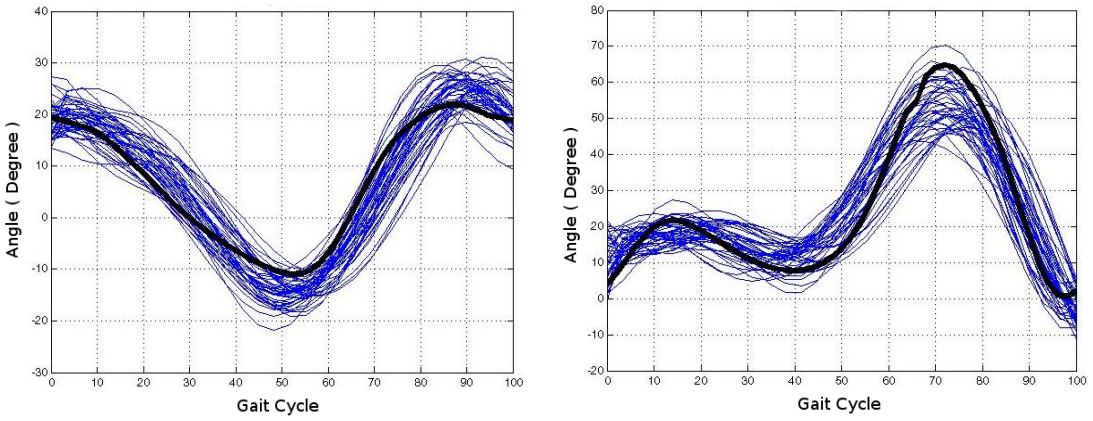
algorithm for 100 subjects are presented in Figure (5.8(b)). It is observed that the average start of the striking phase of the left leg and terminal stance are 24.62% and 50.11% respectively. This is indeed consistent with the medical data [131] as depicted in Figure (3.5).

5.4.2 Joint Extraction

For the evaluation of the model-based method proposed for the extraction of the joints of walking subjects using spatial models, the algorithm is tested on a set of 120 video sequences containing 20 different subjects with 6 sequences for every individual. The videos are taken from the SOTON indoor gait database. The extraction results of the ankle, knee and hip joints are satisfactory for the indoor video sequences with the observation that the localisation of the ankle is more accurate than the hip and knee due to the visibility nature of the ankle. This was also concluded for the case of manual labelling described in Section 3.3.5. The angular measurements for the knee and ankle angles[1] are determined from the joints' trajectories for the 120 video sequences as shown in Figure (5.9). To compare the extracted angular data with the biomechanical data provided by Winter [133], an eight-order polynomial fitting is applied to smooth the angular measurements. Further, it was also confirmed experimentally by [27] to offer

[1]See Figure (3.6) for details of the angles

the best fit for the gait data. The biomechanical data is shown in bold in Figure (5.9). Since the angular motion models have not been used for the derivation of gait features, it can be concluded that the results obtained via this approach are consistent with the biomechanical data.



(a) (b)

FIGURE 5.9: Gait Angular Motion during one Gait Cycle: (a) Hip, (b) Knee

An example of the extraction results for indoor data is shown in Figure (5.10). Furthermore, the model-based method is assessed on 20 different video sequences from the outdoor data taken from the SOTON database. The feature extraction at outdoors is almost as accurate as the indoor database. Figure (5.11) is showing an example of the feature extraction applied on a video sequence from the outdoor database. The joints positions for the knee, hip and ankle are extracted successfully with less accuracy as the indoor data. This is mainly because of the difficulties encountered during the foreground segmentation including shadow and background clutter. For example, at the m^{th} frame, the ankle joint of the right leg is not located correctly due to the shadow on the ground.

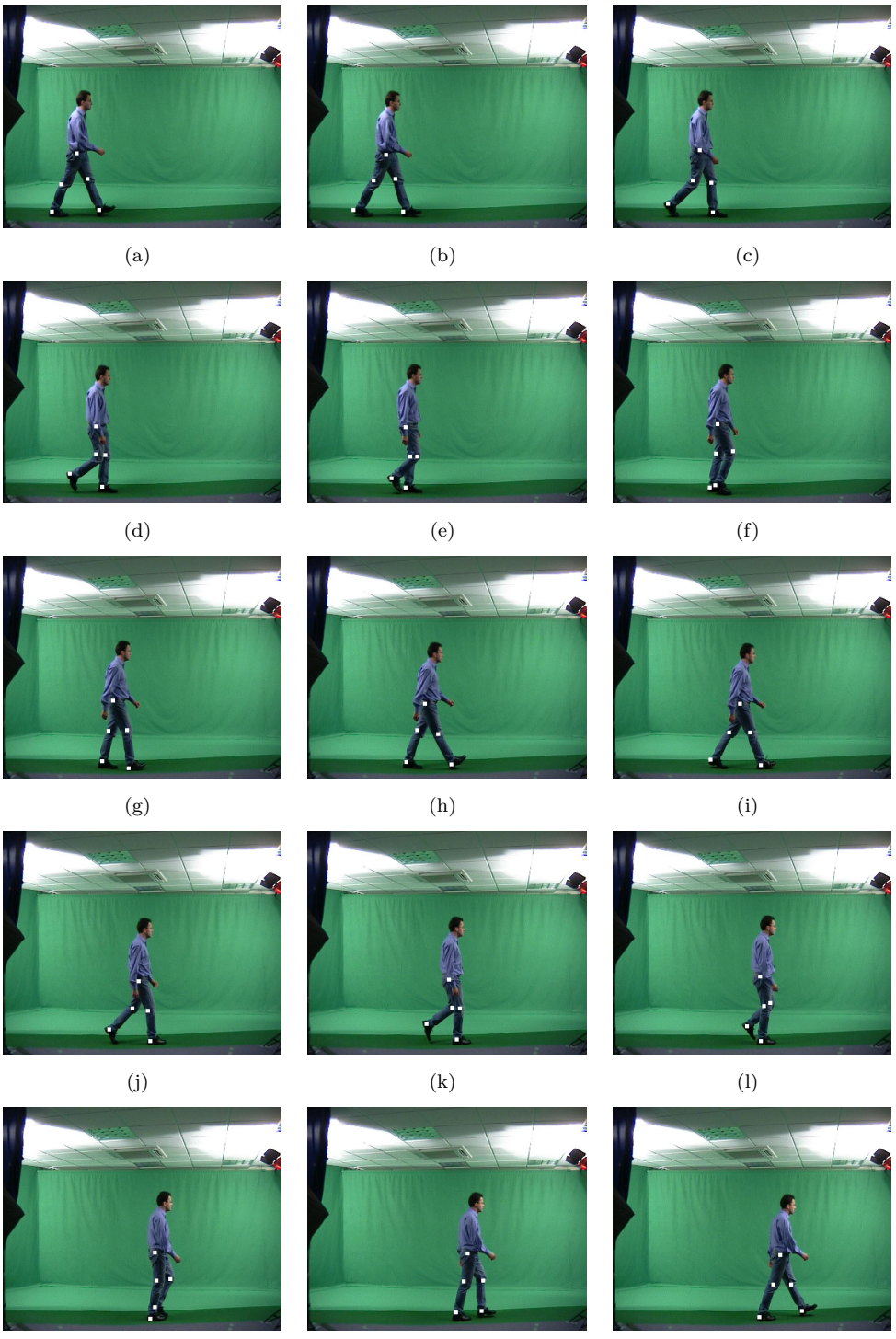


FIGURE 5.10: Automated Extraction of the Joints' Positions for Indoor Data

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

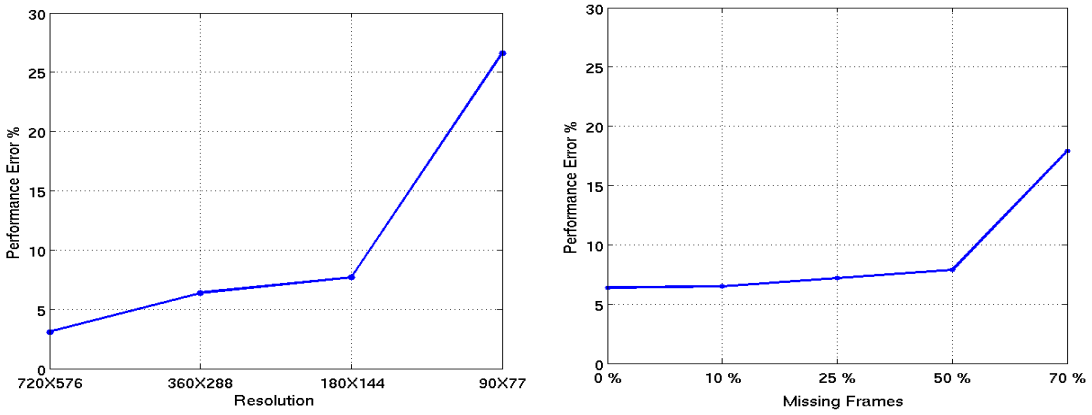
(j) (k) (l)

(m) (n) (o)

FIGURE 5.11: Automated Extraction of the Joints for Outdoor Data

5.4.3 Performance Analysis

The performance of the model-based method proposed for recovering the joints' positions of walking subjects is assessed on video sequences with low resolution. A set of 10 video sequences which are manually labelled, are taken from the SOTON gait database and used for the evaluation process. Figure 5.12(a) presents the performance error of the algorithm for the localisation of the joints at various resolutions. The Euclidean distances between the extracted joints and manually labelled data (i.e., ground truth data) are used to approximate the performance error which is estimated as the average of the distances normalised to the subjects' heights. The resolution of the video sequences are reduced gradually from an original size of 720 × 576 pixels with the aspect ratio kept constant. The algorithm is still able to derive most of the joints with good accuracy at a resolution of 144×180 with a performance error of 7.7%. However, the algorithm fails when the video resolution is reduced to 77×90 pixels.



(a) (b)

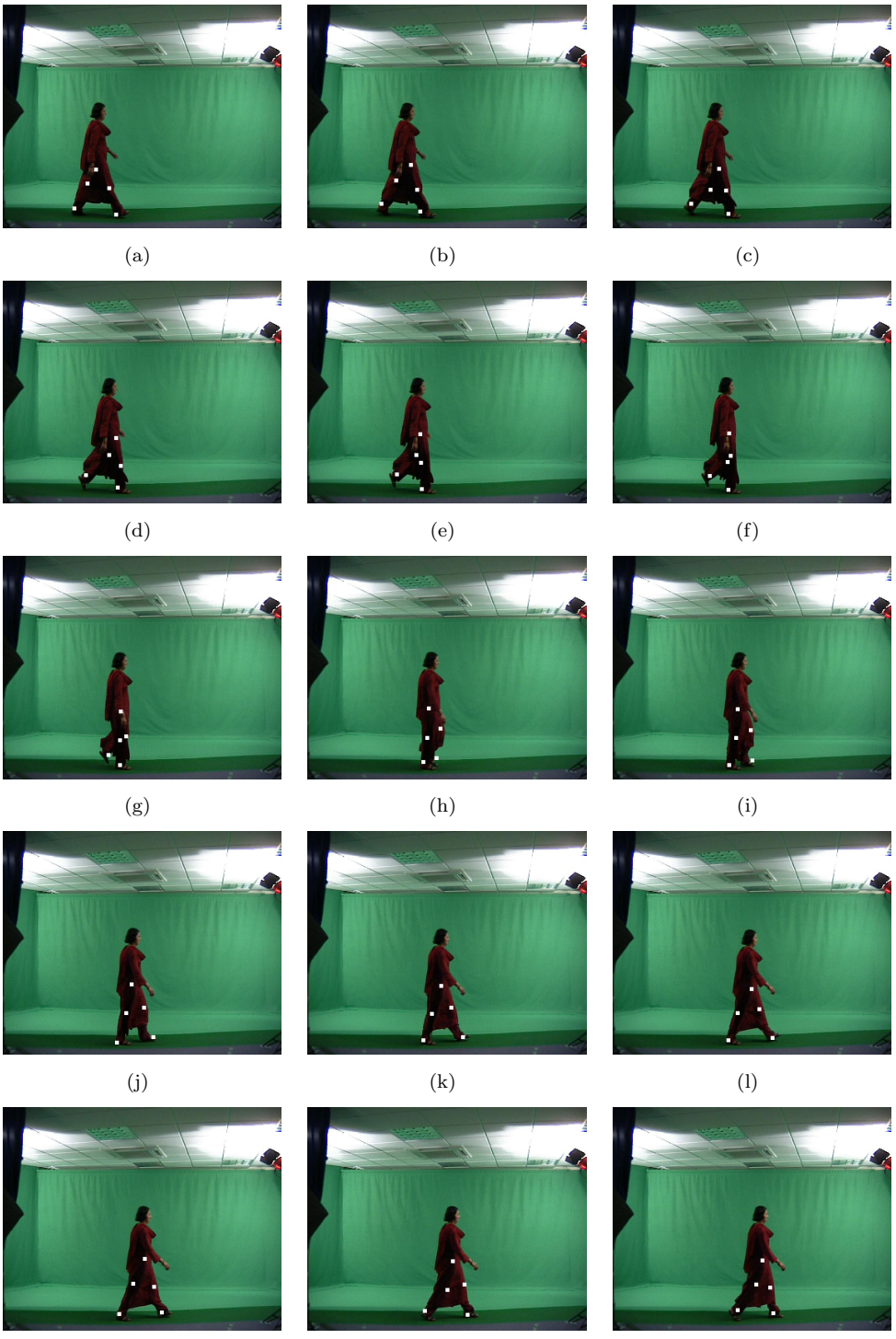
FIGURE 5.12: Performance Analysis for The Joint Extraction.

Furthermore, a number of experiments are carried out using the same video set to investigate the algorithm potentials for handling occlusion. For the case of full occlusion, the performance error is simulated by dropping a number of frames from every 30 frames (30 is the original frame rate) of the video sequences which is equivalent to changing the frame rates. Figure (5.12(b)) shows the performance error estimated by decreasing the frame rates of the video stream resized at resolution of 360x288 pixels. It is observed that

the algorithm performance is not much affected even when dropping 50% of the frames as the algorithm predicts the joint positions for the missing frames using the temporal and spatial models. This proves that one of the merit of model-based methods is their capability of handling occlusion efficiently and recovering the occluded data. Moreover, the algorithm is tested on a subject wearing baggy Indian clothes which covered the legs. The joints positions are extracted successfully as shown in Figure (5.13) which reveals the potentials of this approach to handle also self-occlusion.

5.5 Conclusions

We propose a model-based method to extract moving joints via a recursive evidence gathering technique. Spatial model templates for human motion are derived from the analysis of gait data collected from manual labelling. Model templates are represented in a parametric form based on elliptic Fourier descriptors. Gait knowledge is exploited via heel strike extraction to reduce the parameter space dimensionality and therefore reduce the computational load of the Hough Transform being used in the extraction process. The described method is proved to work for both indoor and outdoor environments with potential to localize joint positions with better accuracy. In the next chapter, recognition potency will be assessed using the data extracted via the described model-based method.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

FIGURE 5.13: The Extraction of Joints for Indoor Data with self-occlusion

Chapter 6

Gait Recognition by Dynamic Cues

The interest in gait as a biometric is strongly motivated by the need for automated recognition systems for visual surveillance and monitoring applications. Most of the methods proposed for gait recognition are model-free methods which rely primarily on silhouette data and statistical methods. Despite the fact that model-free approaches have achieved high recognition rates, these methods were unable to explore and analyse the discriminative power of the different gait features. Since it is possible to extract the joint trajectories using the model-based approach described in the previous chapter, intensive research is conducted to investigate the potential of gait features for people identification using computer vision methods. In this chapter, the different types of features used for gait analysis and recognition are described. The recognition significance of the various dynamic cues which are extracted from the joint angular motion of the lower limbs is examined for constructing a gait signature derived purely from the gait kinematics.

6.1 Static vs. Dynamic Gait Features

Gait features can be broadly classified into two categories, namely static and dynamic cues. The *static features* reflect the geometry-based measurements of the anatomical structure of the human body such as the person's height and the length or width of the different body segments. Static features can also be derived from the observed gait such as the stride length. The *dynamic features* are the cues which describe the kinematics of the locomotion process, such as the angular motion of the lower limbs extracted from the joint trajectory data. As the static cues are less taxing to extract and compute, it would seem straightforward to recognise people using static features such as the stride and body height and so forth. Furthermore, recent research on gait using static features for identification proved that a promising recognition rate can be reached [16, 125]. BenAbdelkader *et al.* [13, 12] demonstrated that gait recognition can be achieved using the subject height and the stride parameters (stride length and cadence) as there is a linear relationship between the two stride variables which can be exploited for recognition.

Some researchers have preferred to fuse both static and dynamic cues with a belief that fusion would yield the optimal gait recognition rate. Wang *et al.* [125] extracted both dynamic and static features of gait motion using a model-based method based on the Condensation framework [57]. Both cues are fused at decision level using different combination of rules such as product and summation. Similarly, Wagg [122] proposed a model-based method for recovering the joint angular motion and the static parameters of the human body such as the thigh width. The f-statistic scores are used as coefficients to weight the discriminative features. Although a high recognition rate was reported, it was clearly noted that the body-related parameters (i.e., static features) are not robust [122] since they are highly dependent on the clothing of walking subjects. Table 6.1 surveys a number of methods which employed either static cues or fusion of static and dynamic features for gait recognition. The symbols LOO and KNN used in the table refer to the Leave-one-out cross-validation and K-nearest neighbour classifier rule.

Method Name	Feature Type	Database Description and Recognition Method	CCR %
BenAbdelkader [12]	Static	45 Subjects with 4 Sequences for every individual. Features include height and stride parameters. *LOO* and *KNN-1* are used for validation.	40
Johnson [64]	Static	54 Sequences for 18 Subjects recorded from the near side view angle. Recognition is computed using Maximum Likelihood	96
Wang [125]	Fused	80 Video sequences for 20 different Subjects. Fusion is applied via summation. Validation: *LOO* with *KNN-1*	96
Wagg [122]	Fused	2163 Gait sequences for 115 subjects from the SOTON database [107]. Features are weighed using f-statistics/ANOVA. Validation: *LOO* with *KNN-1*	84

TABLE 6.1: Gait Recogntion using Static and Fusion of both Static and Dynamic Features.

Despite the fact that static features were proved by recent experiments to achieve promising gait recognition rates, their use for the development of a biometric system is impractical. This is mainly because static features are dependent on clothing [122], bags, and other factors [100, 119] which would certainly affect the recognition performance. On the other hand, based on the visual cues observed from the lights affixed to the human joints, Cutting *et al.* [29] argued that dynamic features contribute significantly more in human recognition than static cues such as height. Cutting concluded that kinematic information is important for gender classification and static features are insufficient to reveal the gender of the walking subject. Moreover, Stevenage *et al.* [112] supported the argument that gait-related cues are more important than body-related features for the purpose of human identification. This conclusion was based on an experiment carried out on 6 different walkers where it was reported that shape information has no effect on the recognition performance.

The answers to the crucial questions of whether kinematic information offers better discriminability than static features and how gait signature constructed purely from gait dynamics is affected by the different covariate factors, will be addressed in this and the following chapters after discussing the different types of gait features and reviewing the recent results of gait recognition using static parameters, bearing in mind that there is

no as yet convincing theoretical framework or theories identifying the motion parameters playing a role in motion-based recognition.

6.2 Gait Identification System

The gait recognition system proposed in this research is a two-stage process consisting of training and testing. In the training stage, the dynamic gait features are extracted using the model-based method from the video sequences of walking people. In order to derive a unique gait signature for every subject, feature selection is applied to search for the best subset of discriminative features. Finally, classification is carried out on the feature vectors of the training data using a supervised pattern recognition method. The gait signatures derived in the classification space are saved in the system database. In the testing or operational mode, new subjects are processed in the same way as in the training stage. The new derived gait signatures are projected into the classification space. In order to recognise or verify the subject's identity using gait, the new signature is compared against all the signatures stored in the system database to find the best match. The training and the testing phases for the gait recognition system are illustrated in the block diagram shown in Figure (6.1).

6.3 Derivation of Gait Signature

The processing and derivation of ultimately discriminative gait features from this trajectory-based data is still an unsolved and challenging problem [126, 76] due to the complexity of the human visual perception system as well as the compound nature of gait motion inherent in the numerous variables associated with it including kinematics, kinetics and anthropometrics [22]. Various different ways of extracting discriminative features from gait sequences have been proposed [89, 90] for the purpose of people identification. An important issue in gait recognition is the derivation of appropriate features that can capture the discriminative individuality from the subject's gait. Such features should respond to crucial criteria such as robustness and invariance to weather conditions,

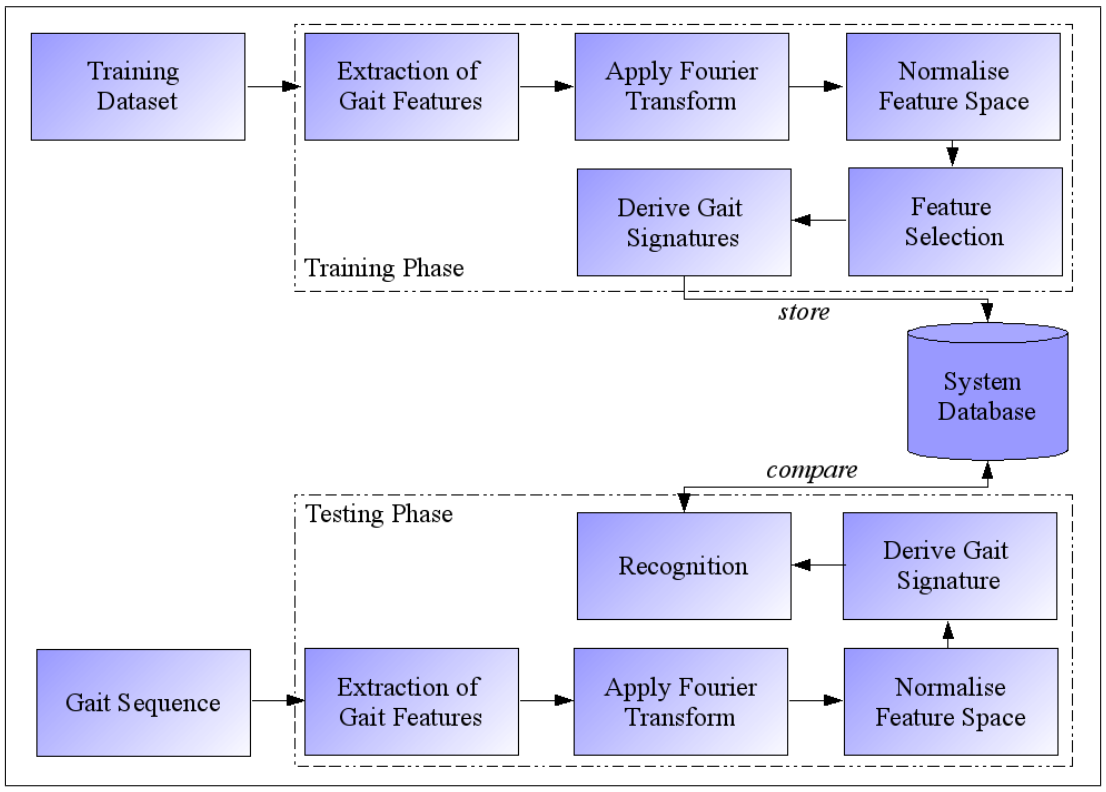


FIGURE 6.1: Structure of Gait Identification System

clothing and operating conditions. They should also yield a good discriminability across different subjects.

6.3.1 Extraction of Dynamic Features

As discussed in earlier sections, psychological studies gave a strong indication that people can use joint angles to recognize human action, classify gender and even to identify walking subjects. Therefore we attempt to develop a human identification system using gait dynamics. The approach is based entirely on the joint angular data that describe the dynamics of the locomotion process. In order to identify a subject by their gait, we need to extract the angular measurements which describe the gait kinematics. The use of angular motion is very common in gait analysis and recognition. The angles of the joints including the hip and the knee are considered the most important kinematics of the lower limbs.

In these studies, the gait vector for each subject is constructed by taking the hip, knee and ankle angular measurements of the right and left legs defined as θ_{rh}, θ_{lh}, θ_{rk} θ_{lk}, θ_{ra} and θ_{la}. The set of symbols: h, k, a, r and l refers to hip, knee, ankle, right and left respectively. It was reported in the medical literature [84] that spatial displacements of the trunk embed some of the discriminative features which reflect the subject's individuality. Therefore, the measurement vector is designed such that it includes both the horizontal and vertical spatial displacements derived from the motion of the hip joint. The displacement values are normalised to the subject height to account for scale-free analysis. The angle and displacement measurements are taken from a single full gait cycle. The gait features can be taken from the averaging over different gait cycles but because of the database limitation, only a single gait cycle is considered.

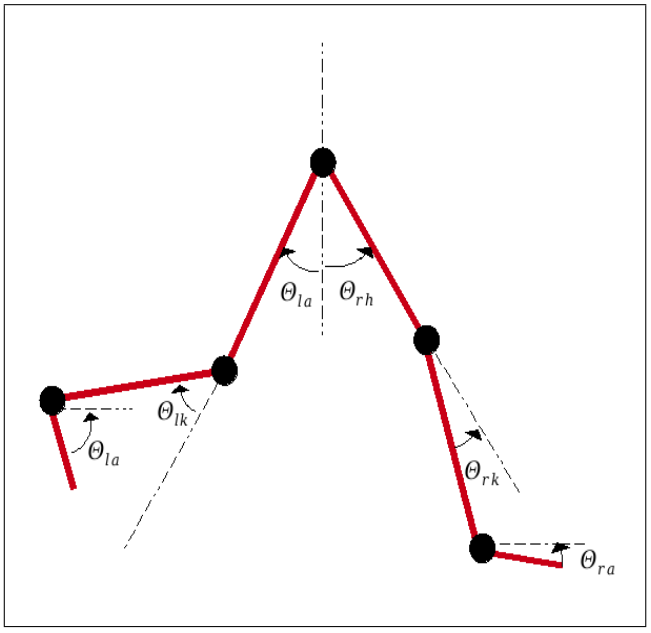


FIGURE 6.2: Gait Angular Measurements.

In order to derive the gait features which embed the relationship of the angular motion between the left and right limbs as well as to improve the recognition rates, additional gait parameters are produced by fusing together the gait angles associated with the hip, knee and ankle. The gait vector is composed by the angle between the thighs called θ_H as the sum of the two hip angles. The vector also includes additional measures produced by combining the right and left angles of the knee as well as the ankle. Combination of

cues is achieved by applying simple rules including summation, product and difference which are denoted as SUM, PRO and DIF producing the following angles: $\theta_s k$, $\theta_p k$, $\theta_d k$, $\theta_s a$, $\theta_p a$ and $\theta_d a$. The size of the angular measurement vector varies from subject to another depending on the length of the gait cycle, but it usually ranges between 24 and 31[1]. To ensure a valid and consistent representation for all subjects in the database, the angular vectors are all resampled to length 32 using cubic spline interpolation.

6.3.2 Fourier Description of Gait Features

In order to facilitate the study and the analysis of the locomotion process and derive the characteristic dynamic features, gait data should be first represented by the basic building blocks because of the complex nature of human gait [131]. One such simplification technique is the Fourier Transform (FT) which transforms complex data and functions into summations of simple sine wave signals, hence greatly simplifying the analysis of gait motion. The FT provides a very efficient and compact gait representation as most of the discriminative information is expected to be contained in a few frequencies. Furthermore, the Fourier Transform enables us to analyse and consider all features without ignoring some of the features as opposed to statistical methods such as Principal Component Analysis (PCA). For each of the normalised gait measurement vectors $(\theta_{rh}, \theta_{lh}, \theta_{rk}, \theta_H, \theta_{lk}, \theta_{sk}, \theta_{pk}, \theta_{dk})$ described in Section (6.3.1), we compute the Discrete Fourier Transform (DFT) for the $N/2$ frequencies of interest in N points.

The gait signature of a walking subject is composed from the magnitude and phase of the Fourier components for the angular data. The phase information has a certain degree of importance in describing the dynamics of the gait pattern. This is because the phase provides the information that describes when the gait dynamics occur. In order to compare phase vectors for different subjects, all analyses are synchronized to start from the same point. This point is chosen as the heel strike of the left leg. Since the magnitude data has been shown to offer low discriminatory capability even though it has the advantages of translation invariance property [26], the product of magnitude

[1]The gait cycle length is measured as the number of frames captured at rate of 25 frames/second.

and phase is used. Hence, the gait signature is constructed as shown in Equation (6.1).

$$f = (Magnitudes \quad Phases \quad Magnitudes \bullet Phases) \tag{6.1}$$

Such that $\bullet$ denotes the element-wise multiplication of magnitude and phase vectors, which weights phase by magnitude to retain proportionate discriminatory ability. The total number of features in f is 675.

6.3.3 Normalisation of Gait Features

In classification, features with large values and high variance may have a greater influence than features with small values. For example, the magnitudes of the Fourier components are significantly larger than the phase values, and as a result the distance between the phases will have little contribution to any magnitude based measure. Henceforth, the efficiency of the classification algorithm will be degraded by the effect of different dimension scaling [50]. This problem can be overcome either by using a different metric measure which takes into consideration the scaling of the different dimensions in the feature space such as the Mahalanobis metric measure, or instead by normalising all the features prior to classification. The normalisation is applied in such a way that all features are linearly scaled so that their values lie in a specific range such as $[0, 1]$:

$$f_n' = \frac{f_n - \min_n}{\max_n - \min_n} \tag{6.2}$$

such that $\min_n$ and $\max_n$ are the minimum and maximum values of the features at the n^{th} dimension.f_n' is the normalized feature. Alternatively, the feature space can be normalised using the z-score [50] which is based on the mean and standard deviation:

$$f_n' = \frac{f_n - \bar{f_n}}{\sigma_{f_n}} \tag{6.3}$$

where $\bar{f_n}$ and σ_{f_n} are the mean and standard deviation of the feature f_n respectively.

6.4 Feature Selection and Classification Metrics

Feature subset selection is the process of choosing the variables that are important for the classification stage from the original feature space. Feature selection is an important task for almost any pattern recognition problem [129]. This procedure is aimed to derive as many discriminative cues as possible whilst removing the redundant and irrelevant information which may degrade the recognition rate. Furthermore, feature selection does not only reduce the cost of recognition by reducing the dimensionality of the feature space, but also offers an improved classification performance through a more stable and compact representation [60, 61]. It is practically infeasible to run an exhaustive search for all the possible combinations of features in order to obtain the optimal subset for recognition due to the high dimensionality of the feature space. For this reason, we employed the Adaptive Sequential Forward Floating Selection (ASFFS) search algorithm[2] [109] which is based on [101]. The Sequential Forward Floating Selection method is experimentally confirmed by Jain *et al.* [141] to perform better than other feature selection algorithms based on experiments carried out for the classification of handprinted characters. An alternative would be to use ANOVA, but this is less favourite in machine learning.

The feature selection procedure fundamentally relies on an evaluation function that determines the usefulness of each feature in order to derive the ideal subset of features for the classification phase. For every feature or set of features generated by the feature selection algorithm, an evaluation criterion is called to measure the discriminative ability of the set of features to distinguish different subjects [30]. A number of methods [83, 138, 118] rely mainly on statistical metric measures which are based on the scatter or distribution of the training samples in the feature space such as the Bhattacharyya metric. These methods aim to find the features which minimize the overlap between the different classes as well as the inner-class scatter.

Although statistical methods enjoy low-cost implementation they have been proved to offer poor estimate of the recognition rate because of their independence from the

[2]The algorithm is implemented by D. Redpath from Heriot-Watt University

final classifier [129, 72, 42]. In this research, the evaluation function is constructed using two criteria: validation-based and statistical-based criteria. The first criterion filters the subsets of features which achieve good recognition rates. The second criterion is applied to the subsets of features chosen by the former function.

6.4.1 Validation-Based Feature Selection

The validation-based evaluation function is proposed to find the subset of features which minimises the classification errors as well as ensure good separability between the different classes. This is achieved when, for every sample in the database, its siblings (i.e., instances of the same class) are the nearest neighbours of such sample. The K-nearest neighbour (KNN) classifier can be employed to offer a good estimate of the classification rate, but it does not ensure that different classes are well separated. Henceforth, a similar measure to the KNN is applied to produce a probability that an instance belongs to the claimed class label based on the nearest neighbours. In contrast to the voting scheme used in the KNN, the scoring function uses different weights w to signify the importance of most nearest neighbours. The probability score for a sample s_c to belong to class c is proposed as:

$$f(s_c) = \frac{\sum_{i=1}^{N_c-1} z_i w_i}{\sum_{i=1}^{N_c-1} w_i} \tag{6.4}$$

where N_c is the number of instances in class c, and the weight w_i for the $i^{th} \in [1, N_c - 1]$ nearest instance is related to proximity as:

$$w_i = (N_c - i)^2 \tag{6.5}$$

The value of z_i is defined as:

$$z_i = \begin{cases} 1 & \text{if } nearest(s_c, i) \in c \\ 0 & \text{otherwise} \end{cases} \tag{6.6}$$

such that the $nearest(s_c, i)$ function returns the i^{th} nearest instance to the sample s_c. The Euclidean distance metric is employed to find the nearest neighbours. The subset

significance based on the validation-based metric is estimated using the leave-one-out cross-validation rule. In the leave-one-out validation, every instance from the original sample is used for testing and is validated against the remaining observations. This is repeated for all the observations in the dataset. The recognition rate is computed as the average of all validations.

6.4.2 Statistical-Based Feature Selection

The statistical evaluation function is based on the scatter of instances in the feature space. It aims to find the set of features which minimises the within-class spread while maximising the between-class spread of data [35]. A number of statistical criteria were proposed for achieving this aim such as in [35, 42]. The system implemented in this research uses a variation of the Bhattacharyya distance measure. The Bhattacharyya distance metric [42] is a measure of the separation score $S_{i,j}$ between class i and j given by:

$$S_{i,j} = (m_i - m_j)\left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1}(m_i - m_j)^T \tag{6.7}$$

such that m_i and Σ_i are the mean and covariance of class i. For the case of N classes, the separation score is computed using the following:

$$J = \frac{1}{N^2}\sum_{a=1}^{N}\sum_{b=1}^{N} S_{a,b} \tag{6.8}$$

The measure J is similar to the decidability measure proposed by Daugman [31] which evaluates how well the within and between class distributions are separated. The decidability value d' is given by:

$$d' = \frac{|\mu_w - \mu_b|}{\sqrt{\frac{\sigma_w^2 + \sigma_b^2}{2}}} \tag{6.9}$$

such that μ_w and μ_b are the means for the within and between class distributions, whilst σ_w and σ_b are their standard deviations. These measures are used later in the analysis.

6.5 Gait Recognition and Performance Analysis

For the evaluation of dynamic gait features derived using the model-based method for people identification, a set of 160 video sequences is taken from the SOTON indoor gait database. The set consists of 20 different subjects walking from left to right with 8 sequence for every individual. There are 10 males and 10 females in the dataset. The selected database serves as a gallery database and is utilised mainly for training the classifier and selecting the optimal feature subset.

6.5.1 Classification Results

Feature selection is applied on the video sequences to search for the most discriminative subset of features. Based on the validation criterion function, the feature selection algorithm derived 34,261 different feature subsets which achieved a recognition rate of 82% based on the probability scores explained in Section (6.4.1). The statistical criteria is applied to choose the best subset of features which offer higher separation between classes. In order to further assess the recognition power of the selected features, the Correct Classification Rate (CCR) is computed using the K-nearest neighbour (KNN) classifier with the leave-one-out cross-validation rule.

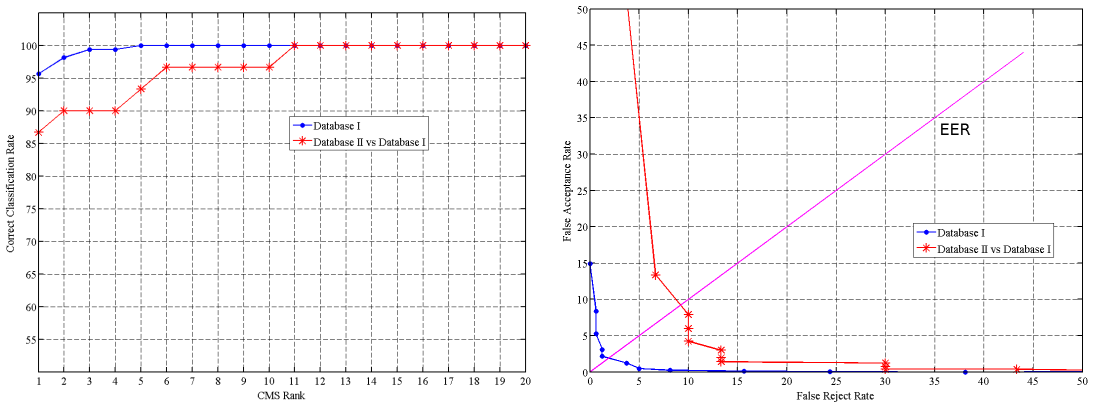
The KNN rule is applied at the classification phase due to its low complexity and hence fast computation besides the ease of comparison to other methods. The KNN classifier uses the Euclidean distance measure to compute the distance between the test and training samples in the feature space in order to find the k closest neighbours based on the Euclidean distances. From the k closest instances, it deduces the class of the test sample by determining the class of the closest neighbours with the highest occurrence frequency. We have achieved a high recognition rate of 95.75% for the value of $k = 5$ using the set of 160 video sequences with 20 different classes. This is achieved using solely features describing purely the dynamics of the locomotion process. The results of the recognition performance are summarized in Table (6.2) with comparative results of other methods which use dynamic parameters for gait recognition.

Method Name	Database and Validation	CCR - *Knn* Rank		
		R=1	R=3	R=5
Wagg [122]	2163 video sequences for 110 subjects, each with 4 sequences. Indoor data, Leave-one-out cross validation is used.	78	-	-
Our method	160 Sequences for 20 subjects from indoors, each with 8 sequences, Leave-one-out cross validation is used.	92.15	93	95.75
Bobick [115]	106 walk sequences for 18 different people. Marker-Based capture system. CCR evaluated using Leave-one-out cross validation	73	-	-
Yam [138]	20 subjects with 5 sequences for every individual. Treadmill Data. Leave-one-out cross validation is used.	84	68	56
Wang [125]	80 video sequences for 20 subjects taken from Indoor data, each with 4 sequences. Leave-one-out cross validation is used.	87.5	-	-
Cunado [27]	10 Test Samples are matched against a database of 30 sequences for 10 subjects. Subjects taken from the SOTON early database	100	100	-

TABLE 6.2: Gait Recognition Results using Dynamic Cues

In order to further assess the classification performance of the proposed method for gait recognition using dynamic-related features, a different dataset which had not been used for feature selection, is taken from the SOTON gait database and matched against the gallery dataset. The dataset is composed of 60 sequences for 20 subjects with 3 sequences for every individual. The first gait dataset (I) which is used for the training and derivation of gait features, is regarded as the gallery such that every instance in the gallery set has its associated class identifier. The second dataset (II) is considered as the probe which consists of samples with unknown class labels. Every sample in the probe database is matched against all the instances in the gallery dataset in order to find the class label for the probe sample using the k-nearest neighbour classifier. For $k = 5$, we have correctly classified 86.67% of the 60 walking sequences. The results achieved using this evaluation are promising because the probe set has not been employed for the derivation of the dynamic feature subset. Henceforth, the derived dynamic features have a potential discriminative capability to identify people.

Another useful evaluation measure is the Cumulative Match Score (CMS) which was introduced by Phillips *et al* in the FERET protocol [99] for the evaluation of face recognition algorithm. The measure assesses the ranking capabilities of the recognition system by producing a list of scores that indicates the probabilities that the correct classification for a given test sample is within the top n matched class labels. In other words, The CMS curve provides a measure of the classification rate of data samples at different ranks or trials to be correctly classified . The CMS curves for the first recognition experiment performed on database (I) as well as the evaluation procedure of database (II) probed against database (I) are depicted in Figure 6.3(a). The scoring function employs the *knn* rule with the value of $k = 5$. It is worth noting that the CMS at rank $R = 1$ is equivalent to the correct classification rate.



(a) (b)

FIGURE 6.3: Classification Results: (a) The Cumulative Match Score Curve. (b) The Receiver Operating Characteristics Curve

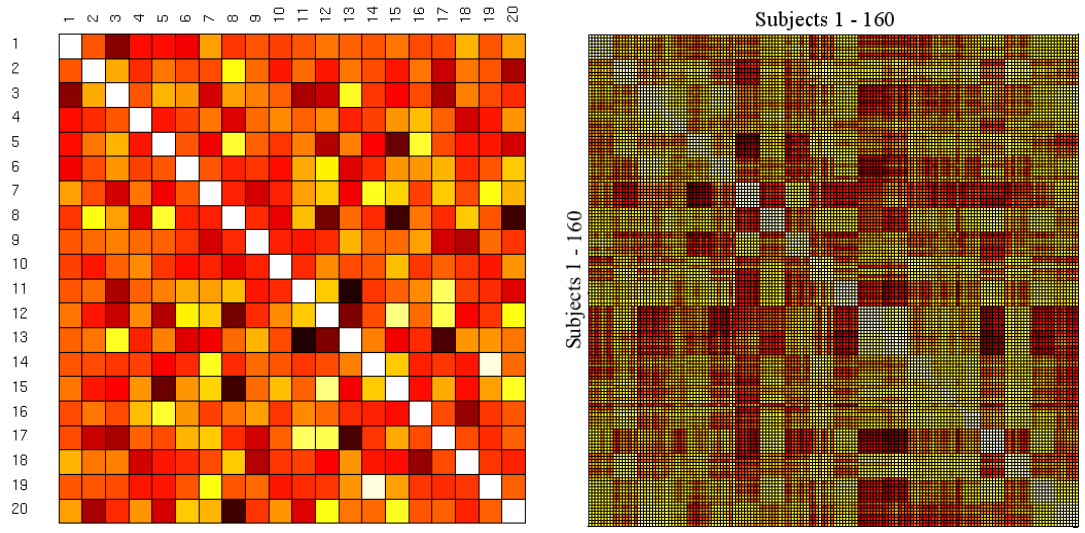
Additionally, The Receiver Operating Characteristics (ROC) curves are plotted in Figure 6.3(b) to express the verification results for the two gait recognition experiments. In the verification process, the instances from database (I) are verified to check if they belong to the claimed class labels based on the nearest neighbours. The probability scoring function described in Section 6.4.1 is used to compute a discrete score based on $k = 5$ nearest neighbours to express whether the sample belong to the claimed class. In order to plot the False Acceptance Rate (FAR) versus the False Rejection Rate (FRR),

different score thresholds are used. Using the gait signature derived from dynamics, the system achieved equal error rates of 3% and 9% for the first and second experiments respectively. To compare our results with other approaches, we have found only two methods which have reported equal error rates. This is because of the lack of work in verification. Bazin *et al.* [11] reported an error rate of 7.3% using fusion of gait features whilst Wang *et al.* [125] achieved a classification error of 8.42% using dynamic gait features.

6.5.2 Statistical Analysis

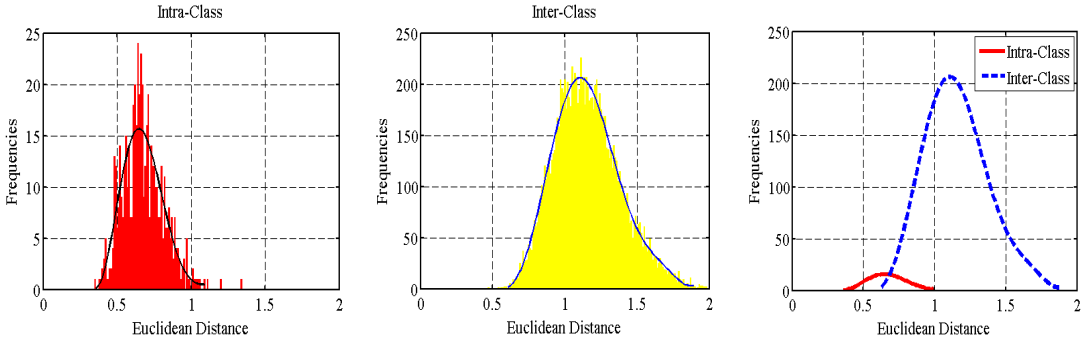
To investigate the discriminatory power of gait recognition using dynamic features, statistical analysis is performed on the gallery dataset (I). The correlation matrix shown in Figure 6.4(a) which visualizes the separation values between the different classes from the first database. The darker squares reflect higher separation score and therefore higher discriminability. The white diagonal line reflects the zero distance between the same class. The separation value between the different classes is estimated using the Bhattacharyya distance metric. Furthermore, all subjects in the database are cross-correlated using the Euclidean metric to produce the correlation matrix of size 160×160 depicted in Figure 6.4(b). The lighter diagonal line represents the comparisons of subjects from the same class.

The distributions of the inter-class and intra-class correlations are shown in Figure (6.5). The intra-class distribution is obtained by comparing all the possible pairings (560 pairs) of the same class. The comparison is based on the Euclidean metric measure and excludes pairs with duplicate items. In the same way, the inter-class distribution is derived by comparing instances which are from different classes (12160 comparisons). The mean and standard deviations for both distributions are summarized in Table 6.3. In order to compare the intra-class and inter-class distributions, two curves are derived from the distributions by fitting 10-order polynomials as shown in Figure (6.5). The distribution representation provides an approximation of how reliably decisions can be made using the chosen distance metric (i.e., Euclidean distance), as the classification



(a) (b)

FIGURE 6.4: Correlation Matrices: (a) Class-Correlation Matrix. (b) Subject-Correlation Matrix



(a) (b) (c)

FIGURE 6.5: Correlation Distributions: (a) Intra-Class Distribution. (b) Inter-Class Distribution. (c) Intra and Inter Class Distributions.

performance is degraded by the overlap of the two distributions [31]. The decidability index described in Section 6.4.2 is estimated from these results as $i = 2.54$.

Distribution	Mean	Standard Deviation
Intra-Class	0.68	0.18
Inter-Class	1.25	0.26

TABLE 6.3: Mean and Standard deviation of Intra and Inter class distributions.

In the evaluation of the recognition performance for dynamic gait features, the k-nearest neighbour rule is used for its simplicity. The KNN is considered as a simple

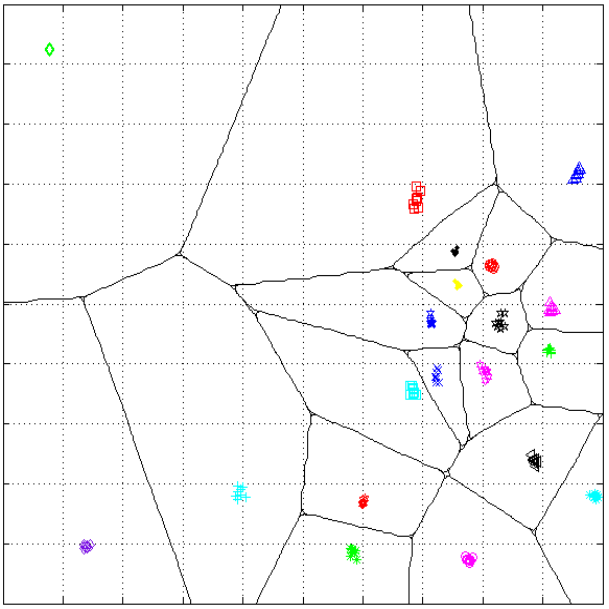


FIGURE 6.6: Projection of Database I using Generalized Discriminant Analysis. The axes are the first two features in the projection space. Every symbol refers to a different subject.

classifier. Therefore, to assess the potential of dynamic-related features, a more sophisticated classifier is employed. The Generalized Discriminant Analysis[3] (GDA) method [10] is applied to optimise the class separability of the first dataset in order to increase the recognition performance. The GDA analysis is a supervised machine learning algorithm derive more discriminative and compact features from the raw features of the training data by applying statistical methods to generation a projection transform. Figure (6.6) shows the projection results of the gait dataset using the GDA Transform. To assess the classification of the projected data in the new space, the leave-one-out cross-validation is used with the KNN rule. For the value of $k = 5$, the system achieved a recognition rate of 100%. The boundaries shown in Figure (6.6) are the boundaries produced by the KNN classifier for every class.

[3]The GDA algorithm used in this research is provided with the Statistical Pattern Recognition Matlab Toolbox implemented by V. Franc from Czech Technical University

6.5.3 Gait Feature Analysis

In order to determine what dynamic features are important for the identification of people, the different gait-related features are analysed separately to estimate their contribution and significance for gait recognition. To produce accurate and unbiased results, we have derived a number of 493 subsets using the validation-based criteria described in Section 6.4.1 for feature analysis. The feature subsets which, are of length between 22 and 54, achieved a Correct Recognition Rate of 92.15% or over using the KNN classifier of $k = 5$. The distributions of the different types of features are illustrated in the pie chart Figure (6.7). The pie-chart provides an indication whether such type of feature is important for the recognition process but it does not provide a measure of its discriminatory potency. Instead, the discriminative significance of dynamic features is approximated using the correct recognition rate based on the KNN classifier and leave-one-out cross validation rule.

Based on the results shown in the distribution Figure 6.7(a) and Figure 6.7(c), it is observed that the angular measurements possess most of the discriminative features with an average proportion of 77% of the gait signature, whilst only a few features are embedded in the displacement motion. the gait angular features are derived from the knee , ankle and hip angles with proportions of 27%, 29% and 21% respectively. The classification results for the dynamic features are plotted in Figure 6.7(b) to measure the discriminatory power of each type of features separately. The knee and ankle angular features are observed to be the most discriminative features with an achieved recognition rate of 52% and 52.5% respectively.

Furthermore, the discriminatory power of the angular measurements versus the displacement features are plotted in Figure (6.7(c)) where it is shown that the combined angular features achieved a recognition rate of 85% whilst only 38.75% is achieved using the displacement features. These analytical results are consistent with the medical reports in [84, 85] whereby Murray observed that the ankle rotation, pelvic tipping and spatial displacements of the trunk embed the subject individuality due to their consistency at different trials. In [121], Wagg has recently confirmed the importance of

the angular features for gait recognition with a reported recognition rate of 77% using dynamic gait features.

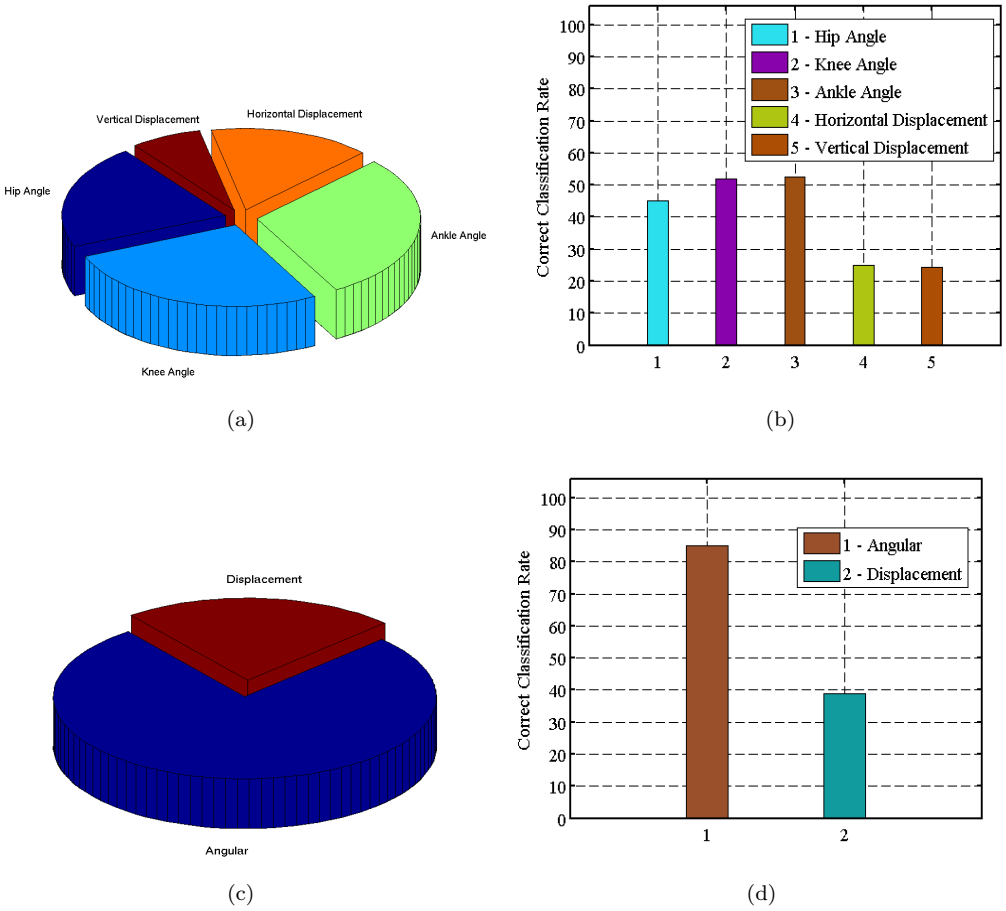
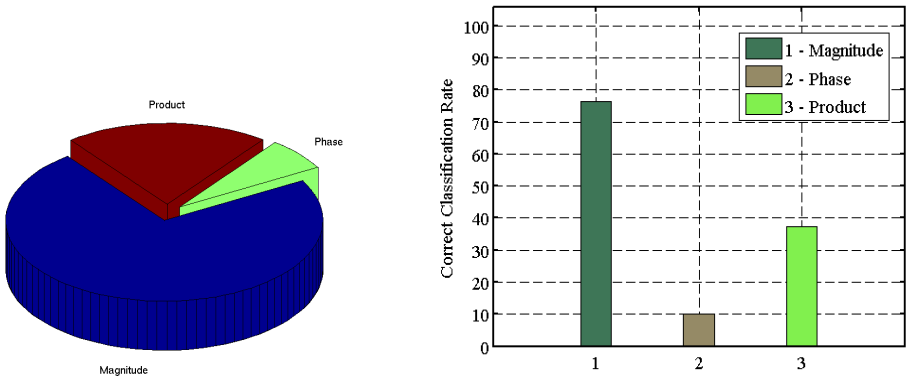


FIGURE 6.7: Dynamic Feature Analysis for Gait Recognition: (a, c) Distribution of Dynamic Features. (b,d) Recognition Performance of Dynamic Features.

The distribution of gait features derived using the Fourier Transform is shown in Figure (6.8(a)). The magnitude of the frequency components is the prominent part of the gait signature with a proportion of 73%. Only a few number of discriminative features are taken from the phase information. The recognition performance of the magnitude versus the phase is plotted in Figure (6.8(b)). A recognition rate of 76.25% is achieved using only the magnitude data, whilst the phase information and Phase Weighted Magnitude (PWM) achieved poor recognition rates of 10% and 37.5% respectively. However with the inclusion of the phase and PWM information into the gait signature, the discriminatory

potency improves to reach 95.75%. This suggests that the phase information is important for the recognition process. The same results are also confirmed by [138, 27, 78].



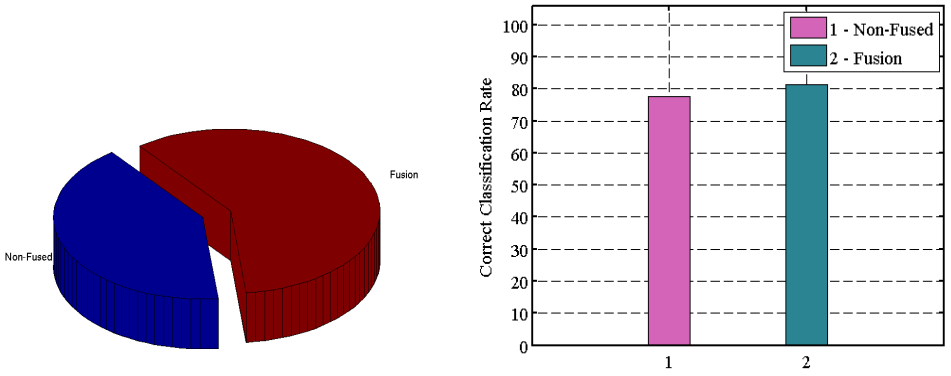
(a) (b)

FIGURE 6.8: Analysis of Fourier Components for Gait Recognitions: (a) Feature Distribution. (a) Gait Recognition Performance.

Figure (6.8) shows the distribution and discriminatory performance of fused features for gait recognition. Most of the gait signature components produced using the Feature Selection algorithm are derived via fusion of dynamic features with a proportion of 59%. For the recognition performance, the fused feature set achieves a correct classification rate of 81%, whilst a non-fused features has a CCR of 77%. This is to conclude that fusion of dynamic features yields more discriminative features which would boost the recognition performance. The same conclusion is confirmed by the work of Bobick *et al* [115] using maker-based solutions for feature extraction.

6.6 Conclusions

In this chapter, results have been reported confirming the early psychological theories claiming that the discriminative features for motion perception and people recognition are embedded in gait kinematics. Dynamic versus static features are reviewed and discussed with their potentials for people identification using gait. We have shown that the gait angular measurements derived from the joint motions mainly the ankle, knee and hip angles, posses most of the discriminatory potency for gait recognition with an



(a) (b)

FIGURE 6.9: Fusion Analysis for Gait Recognition: (a) Feature Distribution. (a) Gait Recognition Performance.

achieved correct classification rate of 95% and an equal error rate (EER) of 3% for verification of people using gait. We have confirmed the usefulness of phase information to improve the classification results of gait recognition as well as the importance role for feature fusion of dynamic features.

Chapter 7

Exploratory Factor Analysis of Gait Recognition

The psychological theories suggesting that people are able recognise each other from the dynamics derived from the joints have been confirmed by the experimental results reported in the previous chapter whereby acceptable recognition rates are achieved using kinematic-related features including angular measures. Nevertheless, such results cannot be generalised to real-world situations due to the fact that subjects are instructed to cooperate with the system without considering the different factors which may affect the performance of gait recognition. Therefore, studying the gait covariate factors becomes a crucial step to translate gait recognition to real-world applications. In this chapter, we review the factors which affect gait recognition using dynamic features including footwear, clothing, carrying conditions and walking speed.

7.1 Covariate Factors in Gait

A non-invasive biometric system for visual surveillance or other security applications, subjects are not required to behave naturally, instructed to walk in a specific way or recorded in a particular environment. For instance, subjects should not be instructed to face the camera (for facial recognition) or walk at normal speed without baggy clothing (for gait identification systems). In fact, there are many variations in the measured data for any biometric type. These variations are called *covariate data* because they are not a fundamental part of the measured property but are a result of human behaviour [92]. The covariate factors can be related either to the subject as for the case when a subject smiles for face recognition, or related to the environmental conditions such as lighting, nature of the ground or camera-setup.

It has been reported by different researchers that one of the main benefit of gait recognition over other biometrics is its non-intrusive nature [123, 12, 53, 89, 92] . Henceforth, the analysis of the covariate factors becomes essential to the understanding of the uniqueness of gait recognition. This will be the focus of this chapter. In fact, the effects of the different covariates for gait analysis and recognition have not been investigated much [92], This is mainly due to the lack of availability of databases, as well as the availability of automated systems which would help for with extraction of gait features. Moreover, the complexity of earlier model-based approaches has precluded their deployment for this analysis. Gait is also affected by different covariate factors including footwear, clothing, injuries, age, walking speed, and much more akin with other biometrics.

The effects of covariate factors on the performance of gait recognition using computer vision methods have been investigated by only one recent major research study by Sarkar *et al.* [105]. Sarker described a baseline algorithm for gait recognition based on the temporal correlation of silhouette data. The algorithm is evaluated on a set of twelve experiments in order to examine the effects of the different covariates including viewpoint, footwear, walking surface, time and carrying conditions. However, their work lacks exploratory analysis of the different gait features under covariate data due to the

use of the silhouette approach. In this research, a full investigation is carried out to explore the covariate effects on gait recognition using dynamic-related features derived via model-based method. The covariate factors includes footwear, clothing, carrying conditions and walking speed. Furthermore, we assess the contribution and discriminatory significance of the different dynamic (gait-related) features used for gait recognition.

7.2 Data Acquisition for Covariate Analysis

7.2.1 Data Acquisition

In order to study the exploratory effects of covariate factors on gait recognition , we have collected a dataset of 440 video sequence from the Southampton Covariate Database described in Section 3.4.2. The dataset consists of ten different walking subjects with eight males and two females. Each subject is recorded from the sagittal view walking at eleven different scenarios, including normal walking. Four video sequences are taken for each situation. The different recorded scenarios are aimed to investigate the influence of the following factors:

- ***Footwear:*** flip-flop, trainer, bare-feet, boots.
- ***Clothing:*** coat, trench coat.
- ***Carrying Conditions:*** barrel bag, handbag.
- ***Walking Speed:*** normal, quick and slow walking.

7.2.2 Gait Feature Extraction

In order to extract the gait features of walking subjects from the covariate dataset, the new model-based method described in Section (5.3) is applied to automate the extraction process of the joint trajectories. Because the main objective of this chapter is to investigate the effects of the different gait covariate factors rather than assess the

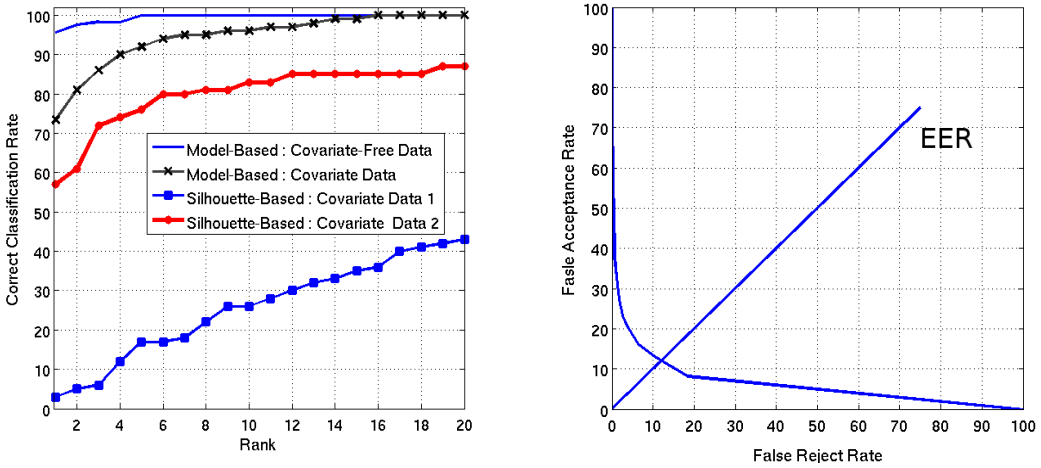
power of the extraction algorithm, the extracted results from all the video sequences are checked and corrected manually where needed to ensure valid and consistent analysis of the covariate effects. The gait signature is composed from the kinematic-related features derived from the measurements of the angular motion and walking displacement. The same methodology described in the previous chapter is used to construct the feature vectors using the optimal discriminative subset of features as well as the classification of gait signatures.

7.3 Covariate Analysis for Gait Recognition

In order to quantity the covariate effects on the performance of gait recognition, feature subset selection is performed to derive a gait signature based on the covariate data as well as the normal gait data. The feature selection process described in the previous gait recognition chapter is purely based on the covariate-free data and therefore may not provide the optimal results for the analysis of the different gait factors. The same gallery dataset of 160 video sequences taken from the SOTON is employed for this study. The gallery consists of 20 different walking subjects with 8 sequences for every individual recorded without covariate effects. The feature subsets which are derived using the validation-based criteria described in Section 6.4.1 are used for covariate-based feature selection. Each of the feature subsets achieved a classification rate of at least 92.15% using leave-one-out validation and KNN classifier applied on the gallery dataset. To derive the optimal subset of features, we use the CCR value computed by probing 440 samples of the covariate data which includes 40 sequences from the normal gait data against the gallery database.

A recognition rate of 73.4% is achieved by validating all the covariate data against the gallery dataset. The covariate factors include footwear, clothing, load carriage and walking speed. The achieved classification rate is higher when compared to the low recognition rates reported by Phillips *et al.* [105] using the silhouette-based method. The Cumulative Match Score curves showing the comparative results are shown in Figure (7.1(a)). The receiver operating characteristic is plotted in Figure (7.1(b)) showing

the verification results for the covariate data with an achieved equal error rate of 13%. Phillips reported a CCR of 57% for data with load carriage and footwear covariates whilst a CCR of 3% is achieved for the following covariates : time, footwear, and clothing. Time changes has been shown [105, 119] to play a major part in reducing recognition capability by gait. Using a silhouette-based approach. Veres showed that this could be redressed by fusing those parts of the gait signature which are invariant with time. In this way the overall CCR could be improved from 23 to 27% [120]. By modelling the feature changes (by using linear interpolation), the recognition rate with variation in time was improved from 23% to 65% [119]. Both of these are considerably improved over the 3% achieved by Sarkar et al [105]. Given the limited data on time, Veres' study and the depth of her results, the time factor is included implicitly and not considered further analysis in this research.



(a) (b)

FIGURE 7.1: Classification Results: (a) The Cumulative Match Score Curve. (b) The Receiver Operating Characteristics Curve

7.3.1 The Footwear Effects

The gait pattern is affected by the different footwear as people are observed to walk differently when wearing trainers as to when wearing flip flops. This has been confirmed by research carried out by Dobbs *et al.* [37]. Based on their experimental results, it

was reported that the stride and cadence parameters of the walking pattern are affected by footwear as opposed to walking with barefeet. Moreover, recent studies [94] showed that changing the footwear texture causes changes in the gait pattern. This is due to the changes in sensory feedback from the surface of the foot [94]. In the studies carried out by Phillips *et al.* [105] to investigate the footwear effects on the performance of gait recognition, a high recognition rate of 78% is reported using a silhouette-based method. This is because of the fact that body-related or silhouette-based features are almost invariant to the different footwear.

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

FIGURE 7.2: Footwear Covariate Factors: (a) Boots (b) Trainer (c) Flip Flops (d) Barefeet

In order to explore the effects of footwear on the performance of people identification using dynamic gait features, a number of experiments are carried out for subjects wearing a variety of different footwear as shown in Figure (7.3.1) including flip flop, boots and normal shoes. In addition, subjects are also recorded walking with barefeet. For each of the footwear-related factors, 40 video sequences are processed to derive gait signatures based on the dynamic gait features. The signatures are taken for 10 different subjects with almost 4 sequences for every individual. To assess the classification performance, subjects are validated against the gallery dataset which consists of 160 gait signatures for 20 different subjects recorded with no covariate effects. The gait signature is composed from the kinematic-related features derived from the measurements of the angular motion and walking displacement. The same approaches described in the Chapters 5 and 6 are used to derive the feature vectors using the optimal discriminative subset of dynamic features.

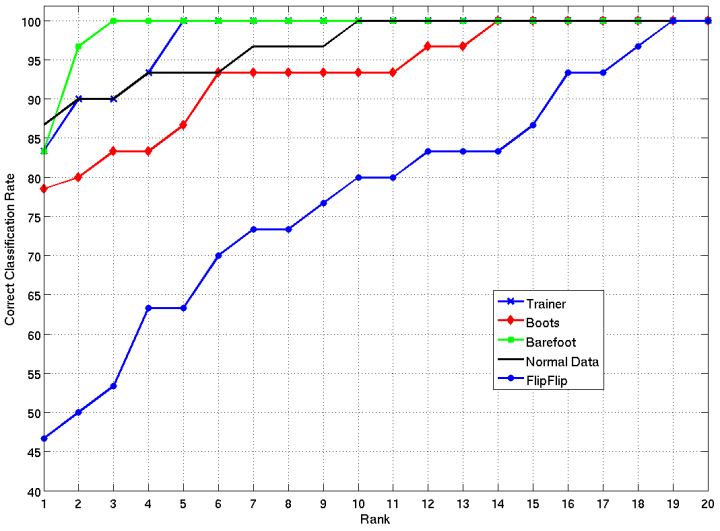


FIGURE 7.3: Classification Results for the Footwear Covariates.

The classification results for the footwear covariates are expressed using the cumulative match score as shown in Figure (7.3). The normal shoe case serves as a reference for other covariate factors. The recognition rates for the trainer and boots cases are observed as high as the normal case with achieved rates of 78%, 83.33% and 86.67% for the boots, trainer and normal shoe cases respectively. When subjects are assessed walking with barefeet, the same gait recognition rate is achieved as the other footwear factors including trainers, boots and normal shoes with a reported CCR of 83.33%. This suggests that the dynamic gait features for people identification are not affected largely with the different footwear. However, the human gait is observed to vary much when people walk with flip flops as the recognition rate drops largely to 46%. The gait features are extracted correctly but the drop of the CCR is likely due to the fact that subjects are not used to wearing flip-flops. Further, there is no rear part of the shoe so this must be compensated when walking.

7.3.2 The Clothing Effects

The clothing effects on human gait as well as the posture and balance can be considerably important. In [102], Punakallio *et al.* showed that suits worn by fire-fighters have significantly impaired their postural and functional balance. In another study by Egan

et al [39], it was revealed that clothing properties such as weight can be another factor which have effects on balance and gait of people. Furthermore, Rahmatalla *et al.* [103] concluded that restrictive clothing can impose constraints on the relative joint angle limits of the walking subject and therefore affect their gait pattern. In the study carried out by Phillips *et al.* [105] for gait recognition using the silhouette-based approach, the recognition rate dropped sharply to 3% for the combined covariate factors time, footwear and clothing.

(a) (b) (c)

FIGURE 7.4: Footwear Covariate Factors: (a) Normal Clothing (b) Coat (c) Trench Coat

In order to investigate the effects of clothing on the human gait and people identification using gait, we have performed a number of experiments on people wearing different clothing including coat, trench coat and their normal clothing as depicted in Figure 7.3.2. For each of the clothing-related factors, the gait signature is derived from the dynamic features for 10 subjects with 4 sequences for every individual. The classification performance is assessed in the same way as the footwear case by matching the probe set against the gallery dataset of 20 subjects with 160 instances recorded with no covariate effects. Figure 7.5 shows the cumulative match score curves for the gait classification experiments. The correct classification rate for the coat is almost the same as the normal case with reported rates of 83.33% and 86.67% respectively. However for the case of the trench coat, the recognition decreases largely to 60%. This is mainly due

to nature of the clothing which is distracting the gait dynamics as well as the occlusion of the knee and hip joints faced during the extraction of gait features. This does not occur with the trousers which adhere to the front of the leg, but could equally occur with the female clothing (e.g., a chardor).

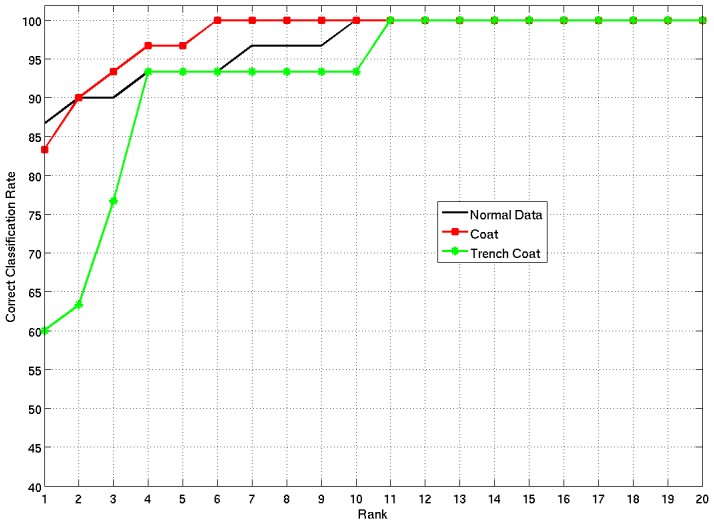


FIGURE 7.5: Classification Results for the Clothing Covariates.

7.3.3 Load Carriage

The impact of load carried on human gait and body posture has been extensively investigated for different purposes including medical, training and military [69] use but rarely for the purpose of gait recognition. In [95], Pascoe *et al.* carried out a number of experiments to examine the effects of carrying bags on gait kinematics for youth people. Pascoe reported that the stride length decreases whilst the gait cadence increases in response to the weight of the load. The same results were also confirmed by the work of Attwells *et al.* [8] and Wang [128]. Attwells observed from experiments carried out on military personnel that the gait angular data including the knee and femur angles are significantly affected with the increase of carriage load. For the effects of carrying conditions on the performance of gait recognition, Phillips *et al.* [105] reported a correct classification rate of 61% using KNN for $k = 1$ employing a silhouette-based method for people carrying a briefcase.

(a) (b) (c)

FIGURE 7.6: Footwear Covariate Factors: (a) Normal Clothing (b) Coat (c) Trench Coat

To investigate the impact of load carriage on the performance of gait recognition using the model-based method for the extraction of dynamic gait features, three different covariate cases related to carrying conditions are used to construct the probe dataset. The cases include people carrying handbags and barrel bags besides the normal walking without carriage as illustrated in Figure (7.6). The probe set consists of 120 video sequences for 10 different subjects with 4 trials for every case. People in the probe set are matched against the same gallery dataset which is used for the evaluation of previous covariate factors. The classification results for gait recognition are detailed using the CMS curves shown in Figure (7.8).

The achieved recognition rate for people carrying a handbag is almost the same as the normal case with a reported CCR of 80%. This is because the handbag was not sufficiently heavy to affect the posture. For the case of the barrel bag which is covering the mid part of the human body and occasionally the legs, the recognition rate drops slightly to 77%. This is because of the occlusion encountered due to the size of the bag and therefore affecting the gait measurements. However, such results may not express the real impact of the carriage load on the performance of gait recognition. This is because the duration of load carriage was brief, as the responses and effects of load may

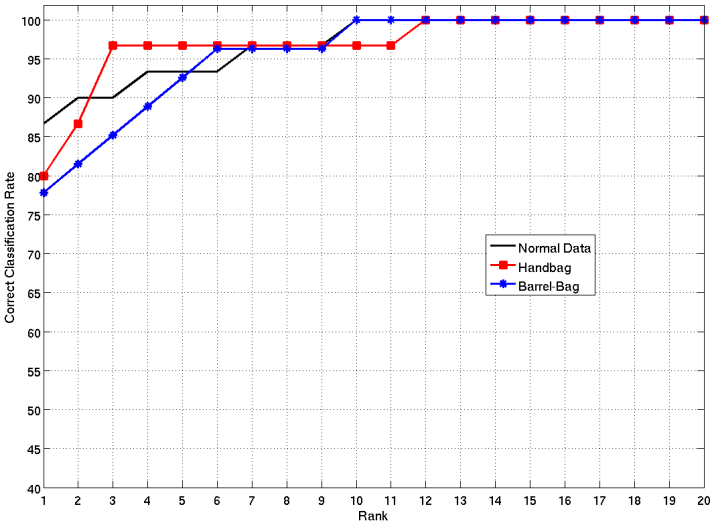


FIGURE 7.7: Classification Results for the Load Carriage.

change with the duration of carriage as a result of fatigue. This was not possible to study in this research due to the limitation of the gait database.

7.3.4 The Speed Effects

There is currently not much work that investigates the effects of speed on the performance of gait recognition methods and the relationship between the gait features and the varying walking speed [116]. Based on a model-based method for feature extraction, Yam [137] reported the possible existence of an individual mapping between the walking and running gait patterns. In [114], Bobick *et al.* observed that appearance-based features derived from silhouette of walking people are speed-dependent and therefore, a preprocessing stage for feature adjustment is suggested to improve the recognition performance. To study the impact of speed variation on gait recognition, a probe dataset is constructed consisting of 10 subjects recorded at different walking speed: slow, normal and quick with 4 trials for every case. The recognition rate for both slow and quick walking drops largely to 60% and 50% respectively compared to the achieved CCR of %86 for the normal walking case.

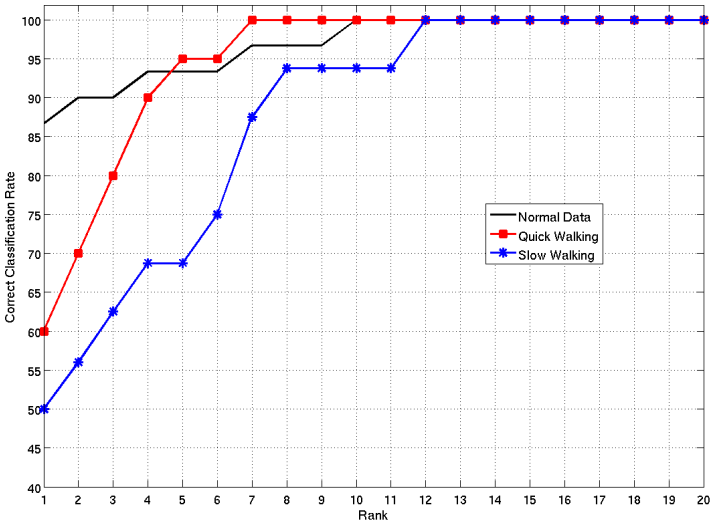


FIGURE 7.8: Classification Results for the Walking Speed Covariates.

7.4 Covariate Factor Analysis of Gait Features

Feature analysis is performed to quantify the footwear effects on the different dynamic gait components employed for recognition. For each of the gait angular signature components (i.e., knee, ankle and hip), the correct recognition rate is computed using leave-one-out validation and a KNN classifier with $k = 5$ for the different covariate factors. The overall results are summarised in Table 7.1 which shows the means and standard deviations of the recognition rates for the various gait dynamic features. The knee is observed to have the highest average CCR whilst it is the most susceptible component to the different covariates with a standard deviation of 14.1%. The ankle has the lowest standard deviation among the angular features. The vertical tipping motion of the trunk (Y displacement) is observed as the most stable features with high average CCR and almost low standard deviation.

TABLE 7.1: Statistical Analysis of Gait Features.

	Hip	Knee	Ankle	X disp	Y disp
Mean	25	27.9	24.1	15.9	23.3
Std. Deviation	12.1	14.1	9.6	7.2	7.3

Figure (7.9) and (7.10) show the results of the angular and displacement feature

analysis respectively for the different covariate factors. The knee and hip angles are less affected when changing the footwear with average CCRs of 34.75% and 31.25% compared the normal case of 37% and 37% respectively. The ankle is observed to have the most impact when changing footwear with an average CCR of 17.25% compared to 37% for the normal case. Interestingly, the ankle angle has almost the same CCR for the other covariate factors including clothing, load carriage and walking speed, whilst the CCR for the knee angle drops largely for the clothing covariates. The features derived from the horizontal and vertical displacement of the trunk are shown to have less impact from the covariate factors in contrast to the angular components of the gait signature. Performance analysis on the effects of the covariates combined together could not be done in this research due to the limitation of the database.

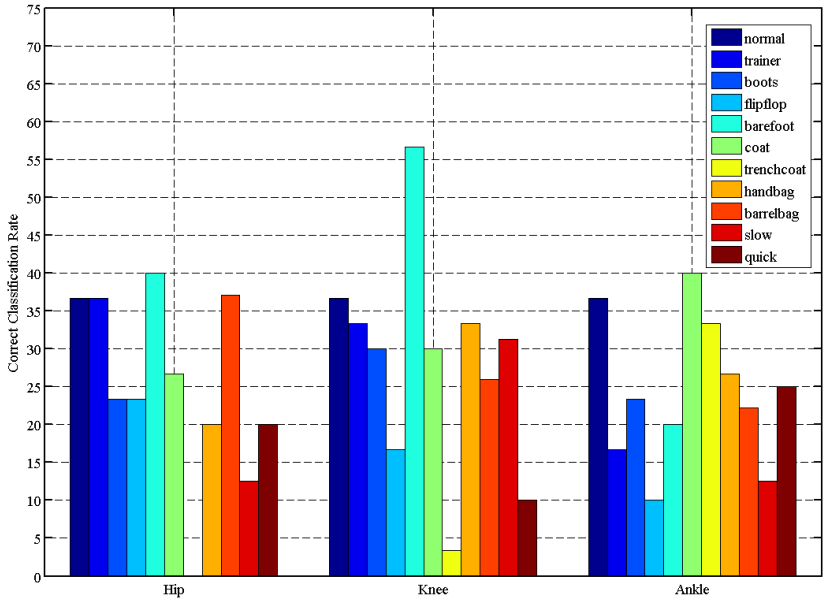


FIGURE 7.9: Angular Feature Analysis for Covariate Gait Recognition.

7.5 Conclusions

In this chapter, we have investigated the impact of the different covariate factors on the performance of gait recognition using kinematic-related features. Four different covariates are analysed including footwear, load carriage, clothing and walking speed.

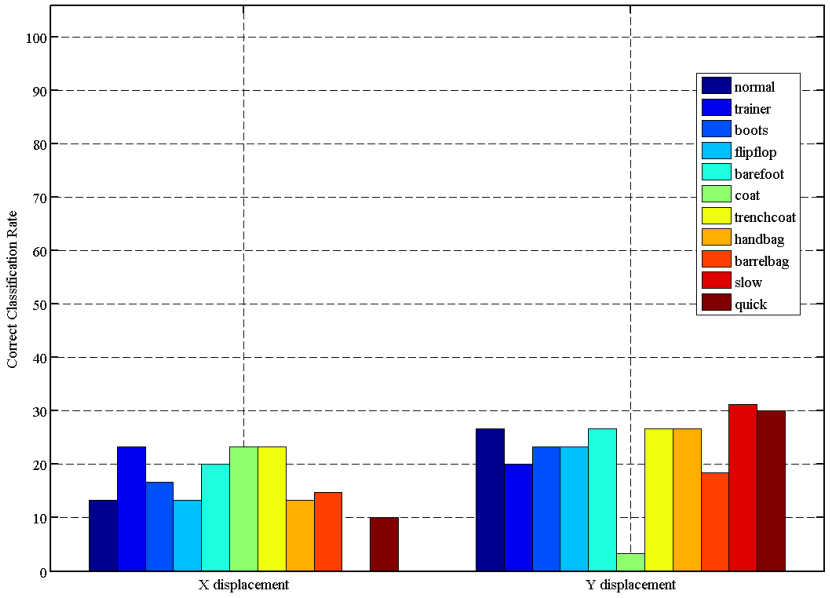


FIGURE 7.10: Displacement Feature Analysis for Covariate Gait Recognition.

Based on a covariate-based probe dataset of 440 samples, a high recognition rate of 73.4% is achieved using the KNN classifier with $k = 5$. This is to conclude that people identification using dynamic gait features is still perceivable with better recognition rate even under the different covariate factors. The footwear, clothing and load carriage covariates are observed to have almost no effects on the performance of gait recognition with similar results when walking with barefeet or without carrying bags. However, the gait recognition drops largely when walking with flip flops or wearing a trench coat due the difficulties encountered during the extraction of dynamic gait features using the model-based method. The difficulty of extracting obscured features pervades computer vision; if the features are completely obscured, it is unlikely that an algorithm can be designed to successfully extract them.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This thesis has investigated the use of kinematic-related gait features for people detection and identification as well as the possibility of translating the use of gait recognition to visual surveillance applications. We have proposed a model-based method for the recovery of the joint trajectories. This is because the model-based approach is more suited to general deployment including cases where subjects are captured walking from different viewpoints. The main thrust behind this research is the complementary studies from psychology and various other disciplines which supported the founding concept that gait is unique for every person and people can be recognised by the way they walk.

A new approach was described to classify moving objects and detect walking people. Multiple objects are tracked successfully through the use of shape-based parameters to allocate them to different layers. Problems encountered during tracking such as background clutter, appearance of uninteresting objects and entry and exit of objects are handled efficiently. In contrast to approaches that employ shape-based parameters for classification, we have explored an alternative technique for walking people detection based on their gait motion. The rhythmic pattern of the gait motion derived from the heel strike pattern is utilised as the main cue to distinguish walking subjects from

other moving objects. The experimental results confirm the robustness of our method to discriminate between a single walking person, a group of people and a vehicle with a detection rate of 100%. Furthermore, the algorithm proposed for the detection of the heel strikes is proven to derive the strikes with high accuracy for different type of video streams including low resolution of 72×90.

We have proposed a new model-based method to extract moving joints via a recursive evidence gathering technique. The method is based on the spatial model templates describing the human motion. The motion models are derived from the analysis of gait data collected from manual labelling and are represented in a parametric form based on Elliptic Fourier Descriptors. Because the parametric representation of the motion models requires 5 parameters and henceforth it is computationally infeasible to implement, the gait knowledge is exploited via heel strike extraction to reduce the parameter space dimensionality and reduce the computational load of the Hough Transform being used in the extraction process. The described method is shown to work for both indoor and outdoor environments with potential to localize joint positions with better accuracy.

We have carried out experiments to confirm psychological claims that the discriminative features for motion perception and people recognition are embedded in gait kinematics. The gait signature for people identification is constructed using a validation and statistical-based feature selection process. Based on features describing the pure dynamics of gait which are derived from the angular measurements of the lower limbs as well as displacement of the trunk, a high correct classification rate of 95% and an equal error rate (EER) of 3% for verification of people using gait are reported with the possibility of improving these rates using sophisticated classifiers. We have confirmed using computer vision methods, that the ankle, knee and hip rotation possess most of the discriminatory potency for gait recognition as reported in the medical literature. We have also confirmed the usefulness of phase information to improve the classification results of gait recognition as well as the importance role for feature fusion of dynamic features.

We have investigated the impact of the different covariate factors on the performance

of gait recognition using kinematic-related features. The covariates include footwear, load carriage, clothing and walking speed. Based on a covariate-based probe dataset of 440 samples, a high recognition rate of 73% is achieved using the KNN classifier compared to a CCR of 57% using a silhouette-based method by Sarkar *et al.* [105] applied on data with footwear, clothing and time covariates. This is to conclude that people identification using dynamic gait features is still perceivable even under the different covariate factors.

8.2 Future Work

Although the marker-less model-based method presented for the extraction of gait features has shown its effectiveness for different types of walking scenarios, the method has been tested only on data of walking people recorded from the sagittal view. Therefore, further work is required to address the different viewpoint angles for gait recognition. For the recovery of the angular measurements and deriving view-invariant gait signature from the different viewing planes into the sagittal plane, the approach described by Spencer *et al.* [110] should be considered for this research. Furthermore, the performance of the bottom-up model-based approach which is satisfactory for extracting the most important dynamic gait features for people recognition, needs to be assessed further and compared with other top-down model-based approaches [27, 138, 122].

In addition, as the importance of the dynamic gait features for people identification is confirmed in this study, further research should be carried out to investigate the discriminatory power and analyse the kinematic characteristics of gait motion using more advanced statistical methods in order to derive more discriminative and efficient features from the gait dynamics. Although the achieved results for gait recognition are encouraging using only the angular and displacement measurements, the classification performance could be improved further by using more features either derived from the anthropometric measurements of the human body (static features such as height and leg width) or derived using more advanced fusion techniques.

Last but not least, due to the limitation of the gait database, we have only investigated the impact of four covariates on the performance of gait recognition including footwear, load carriage, walking speed and clothing. We could extend this analytical studies to assess the effects of other covariate factors including walking surface, viewpoint and more importantly analyse the impact of time factor on the model-based approach. In addition, further research is needed to classify and detect the different covariates or walking modes based on the derived features as well as developing a mapping algorithm to tune or adjust the gait features according to the detected covariate factor or walking mode. Henceforth, the gait recognition rate could be improved significantly by exploiting such mapping between the gait features and covariate factors. As such a program of research has addressed many of the powerful factors in automatic analysis of surveillance video, the abstraction and analysis of human movement from within surveillance video remained to be performed and we look forward to the contribution of the techniques developed during this thesis to those developments.

Appendix A

Parameters Reduction

Let H the curve we are searching for using the recursive evidence gathering algorithm and, s_i is the i^{th} heel strike point. Intuitively, H is derived from model template F by applying appearance transformations. By incorporating gait knowledge, The curve H is enclosed between the two consecutive strikes of the same leg (i.e. s_i and s_{i+2}). If e_1 and e_2 are the end points of the the ankle joint for the model template F as shown in Figure (5.4(a)), then e_1 and e_2 are mapped to s_i and s_{i+2} on the curve H respectively for the case of the ankle, whilst mapped to different points whose x coordinates are the same as the striking points s_i and s_{i+2} for the case of the knee and hip joints, the y coordinates are approximated to be the same as shown in Figure (5.4(a)). The points e_1 and e_2 are expressed in their complex form:

$$\begin{cases} e_1 = x_{e1} + iy_{e1} \\ e_2 = x_{e2} + iy_{e1} \end{cases} \tag{A.1}$$

Then by scaling the template F by s_x and s_y through the horizontal and vertical axis respectively, both points e_1 and e_2 will be shifted to $e_1^{'}$ and $e_2^{'}$ which are defined as:

$$\begin{cases} e_1^{'} = s_x x_{e1} + s_y y_{e1} i \\ e_2^{'} = s_x x_{e2} + s_y y_{e2} i \end{cases} \tag{A.2}$$

For proper scaling, the length of the vector $\overrightarrow{e_1^{'}e_2^{'}}$ should have the same length as the vector $s_i\overrightarrow{s_{i+2}}$ i.e. distance between the two strike points s_i and s_{i+2}. This is illustrated in equation (A.3):

$$| \overrightarrow{e_1^{'}e_2^{'}} | = | s_i\overrightarrow{s_{i+2}} | \tag{A.3}$$

Which is equivalent to:

$$| \overrightarrow{e_1^{'}e_2^{'}} | = \sqrt{s_x^2(x_{e2} - x_{e1}))^2} \tag{A.4}$$

Therefore the value of the parameter s_x can be computed using the following equation:

$$s_x = \frac{| s_i\overrightarrow{s_{i+2}} |}{\sqrt{(x_{e2} - x_{e1})^2}} \tag{A.5}$$

The rotation transform of angle α is then applied to the points $e_1^{'}$ and $e_2^{'}$ to derive new mapping points $e_1^{''}$ and $e_2^{''}$ defined in the following equation (A.6):

$$\begin{cases} e_1^{''} = (\cos(\alpha) + i\sin(\alpha))(s_x x_{e1} + s_y y_{e1} i) \\ e_2^{''} = (\cos(\alpha) + i\sin(\alpha))(s_x x_{e2} + s_y y_{e2} i) \end{cases} \tag{A.6}$$

The angle of the vector $\overrightarrow{e_1^{''}e_2^{''}}$ must be equal to the angle value of $s_i\overrightarrow{s_{i+2}}$ as expressed in equation (A.7):

$$angle(s_i\overrightarrow{s_{i+2}}) = angle(\overrightarrow{e_1^{''}e_2^{''}}) \tag{A.7}$$

By substituting:

$$angle(s_i\overrightarrow{s_{i+2}}) = angle((cos(\alpha) + \sin(\alpha))is_x(x_{e2} - x_{e1})) \tag{A.8}$$

Which is equivalent to:

$$angle(s_i\overrightarrow{s_{i+2}}) = angle(s_x(x_{e2} - x_{e1})\cos(\alpha) + i\sin(\alpha)) \tag{A.9}$$

Because:

$$\alpha = angle(s_x(x_{e2} - x_{e1})\cos(\alpha) + i\sin(\alpha)) \tag{A.10}$$

Hencefore, the value the parameter α can be deduced as presented in equation (A.11)

$$\alpha = angle(s_i\overrightarrow{s_{i+2}}) \tag{A.11}$$

By simplifying the parameter equation (5.6), then for a given point p of coordinates (x, y) from the motion pattern H, there exists $t \in [0, 2\pi]$ value such that:

$$\begin{cases} x = x_0 + X(t)s_x\cos(\alpha) - s_yY(t)\sin(\alpha) \\ y = y_0 + s_yY(t)\cos(\alpha) + s_xX(t)\sin(\alpha) \end{cases} \tag{A.12}$$

Let the point $m(x_m, y_m)$ the mapping of the end point e_1 from the model template, then the point m is defined as::

$$\begin{cases} x_m = x_0 + x_{e1}s_x\cos(\alpha) - s_yy_{e1}\sin(\alpha) \\ y_m = y_0 + y_{e1}s_y\cos(\alpha) + s_xx_{e1}\sin(\alpha) \end{cases} \tag{A.13}$$

By subtracting equation (A.13) from (A.13):

$$(y - y_m) = s_y(Y(t) - y_{e1})\cos(\alpha) + s_x(X(t) - x_{e1})\sin(\alpha) \tag{A.14}$$

Therefore, the parameter s_y can rewritten as:

$$s_y = \frac{(y - y_m) - s_x(X(t) - x_{e1})\sin(\alpha)}{(Y(t) - y_{e1})\cos(\alpha)} \tag{A.15}$$

For the global search of the ankle motion pattern, the point m is the striking point s_i. Thus, the dimensionality of parameter space is reduced to one parameter s_y as y_m is equal to y coordinate of s_i. For the case of the knee and hip joints, the two parameters s_y and y_m are used during the search.The translation parameters x_0 and y_0 are then computed after determining the best value of s_y and y_m using equation (A.13):

Appendix B

Anthrometric Measurements of the Human Body

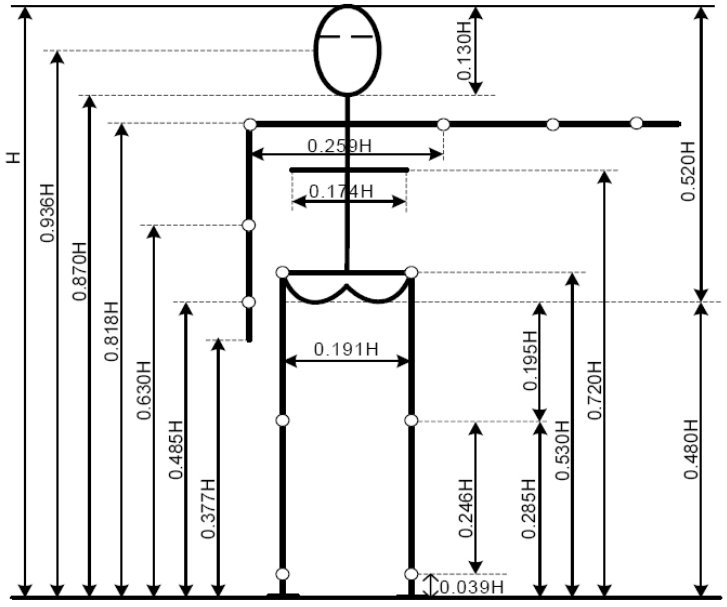


FIGURE B.1: Body segment properties [132].

References

[1] J. K. Aggarwal. Problems, Ongoing Research and Future Directions in Motion Research. *Machine Vision and Applications*, 14(4):199–201, 2003.

[2] J. K. Aggarwal and Q. Cai. Human Motion Analysis: a Review. *Computer Vision and Image Understanding*, 73(3):428–440, 1999.

[3] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Nonrigid Motion Analysis: Articulated and Elastic Motion. *Computer Vision and Image Understanding*, 70(2):142–156, 1998.

[4] J. K. Aggarwal, L. S. Davis, and W. N. Martin. Correspondence Processes in Dynamic Scene Analysis. *In Proceedings of the IEEE*, 69(5):562–572, 1981.

[5] A. S. V. Aguado, E. V. Montiel, and M. S. Nixon. Bias Error Analysis of the Generalised Hough Transform. *Journal of Mathematical Imaging and Vision*, 12(1):25–42, 2000.

[6] A. S. V. Aguado, M. S. Nixon, and M. E. Montiel. Parameterizing Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction. *Computer Vision and Image Understanding*, 69(2):202–221, 1998.

[7] K. Akita. Image Sequence Analysis of Real World Human Motion. *Pattern Recognition.*, 17(1):73–83, 1984.

[8] R. Attwells, S. Birrell, R. Hooper, and N. Mansfield. Influence of Carrying Heavy Loads on Soldiers' Posture, Movements and Gait. *Ergonomics*, 49(14):1527–1537, 2006.

[9] D. H. Ballard. *Generalizing the Hough Transform to Detect Arbitrary Shapes.* Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1987.

[10] G. Baudat and F. Anouar. Generalized Discriminant Analysis using a Kernel Approach. *Neural computation*, 12(10):2385–2404, 2000.

[11] A. I. Bazin and M. S. Nixon. Gait Verification Using Probabilistic Methods. *Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05)*, 1:60–65, 2005.

[12] C. BenAbdelkader, R. Cutler, and L. Davis. Person Identification using Automatic Height and Stride Estimation. *In Procedings of the International Conference on Pattern Recognition*, 4:377–380, 2002.

[13] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and Cadence as a Biometric in Automatic Person Identification and Verification. *In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 357–362, 2002.

[14] G. P. Bingham, R. C. Schmidt, and L. D. Rosenblum. Dynamics and the Orientation of Kinematic Forms in Visual Event Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1473–93, 1995.

[15] R. Blake and M. Shiffrar. Perception of Human Motion. *Annu. Rev. Psychol*, 58:12–1, 2007.

[16] A. F. Bobick and A. Johnson. Gait Recognition using Static Activity-Specific Parameters. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:423–430, 2001.

[17] B. Bodenheimer, C. Rose, S. Rosenthal, and J. Pella. The Process of Motion Capture: Dealing with the Data. *Computer Animation and Simulation*, 97:3–18, 1997.

[18] I. Bouchrika and M. S. Nixon. Model-Based Feature Extraction for Gait Analysis and Recognition. *In Proceedings of Mirage: Computer Vision / Computer Graphics Collaboration Techniques and Applications*, pages 150–160, 2007.

[19] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis. Gait Recognition: a Challenging Signal Processing Technology for Biometric Identification. *IEEE Signal Processing Magazine*, 22(6):78–90, 2005.

[20] L. W. Campbell and A. F. Bobick. Recognition of Human Body Motion using Phase Space Constraints. *In Proceedings of the Fifth International Conference on Computer Vision*, pages 624–630, 1995.

[21] C. Cedras and M. Shah. Motion-based Recognition: A survey. *Image and Vision Computing*, 13(2):129–155, 1995.

[22] T. Chau. A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. *Gait Posture*, 13(1):49–66, 2001.

[23] F. S. Chen, C. M. Fu, and C. L. Huang. Hand Gesture Recognition using a Real-Time Tracking Method and Hidden Markov Models. *Image and Vision Computing*, 21(8):745–758, 2003.

[24] D. Comaniciu, V. Ramesh, and P. Meer. Real-time Tracking of Non-Rigid Objects using Mean Shift. *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition*, 2, 2000.

[25] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting Moving Objects, Ghosts, and Shadows in Video Streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, 2003.

[26] D. Cunado, M. S. Nixon, and J. Carter. Using gait as a Biometric, via Phase-Weighted Magnitude Spectra. *In Proceedings of 1st International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 95–102, 1997.

[27] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic Extraction and Description of Human Gait Models for Recognition Purposes. *Computer Vision and Image Understanding*, 90(1):1–41, 2003.

[28] R. Cutler and L. S. Davis. Robust Real-Time Periodic Motion Detection, Analysis, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781–796, 2000.

[29] J. E. Cutting and D. Proffitt. Gait Perception as an Example of How we may Perceive Events. *Intersensory perception and sensory integration*, pages 249–273, 1981.

[30] M. Dash and H. Liu. Feature Selection for Classification. *Intelligent Data Analysis*, 1(3):131–156, 1997.

[31] J. G. Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, 1993.

[32] W. T. Dempster and G. R. L. Gaughran. Properties of Body Segments Based on Size and Weight. *American Journal of Anatomy*, 120(1):33–54, 1967.

[33] A. M. Derrington, H. A. Allen, and L. S. Delicato. Visual Mechanisms of Motion Analysis and Motion Perception. *Annual Review of Psychology*, 55(1):181–205, 2004.

[34] J. Deutscher, A. Blake, and I. Reid. Articulated Body Motion Capture by Annealed Particle Filtering. *Proceedings IEEE International Conference on Computer Vision Pattern Recognition*, 2:126–133, 2000.

[35] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, New Jersey etc, 1982.

[36] W. H. Dittrich. Action Categories and the Perception of Biological Motion. *Perception*, 22(1):15–22, 1993.

[37] R. J. Dobbs, A. Charlett, S. G. Bowes, C. J. A. O'neill, C. Weller, J. Hughes, and S. M. Dobbs. Is This Walk Normal? *Age and Ageing*, 22(1):27, 1993.

[38] R. O. Duda and P. E. Hart. Use of the Hough Transformation to detect Lines and Curves in Pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[39] W. E. Egan, D. P. Fisher, L. D. Gerber, B. S. Hatler, G. P. Ernst, C. Hall, and N. Henderson. Postural Sway of Subjects Wearing the US Army Chemical Protective Ensemble after Functional Activity. *Aviat Space Environ Med*, 72(9):831–5, 2001.

[40] D. Fish and C. S. Kosta. Walking Impediments and Gait Inefficiencies in the CVA Patient. *Journal of Prosthetics and Orthotics*, 11(2):33–37, 1999.

[41] J. P. Foster, M. S. Nixon, and A. Prügel-Bennett. Automatic Gait Recognition using Area-Based Metrics. *Pattern Recognition Letters*, 24(14):2489–2497, 2003.

[42] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, 1990.

[43] D. M. Gavrila. Visual Analysis of Human Movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.

[44] M. Gleicher. Animation from Observation: Motion Capture and Motion Editing. *Computer Graphics*, 4(33):51–55, 1999.

[45] N. H. Goddard. *The Perception of Articulated Motion: Recognizing Moving Light Displays.* PhD thesis, University of Rochester, 1992.

[46] N. H. Goddard. Incremental Model-Based Discrimination of Articulated Movement from Motion Features. *In Proceedings of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, pages 89–95, 1994.

[47] G. H. Granlund. Fourier Preprocessing for Hand Print Character Recognition. *IEEE Transactions on Computers*, 21(2):195–201, 1972.

[48] M. G. Grant, M. S. Nixon, and P. H. Lewis. Extracting Moving Shapes by Evidence Gathering. *Pattern Recognition*, 35(5):1099–1114, 2002.

[49] Y. Guo, G. Xu, and S. Tsuji. Understanding Human Motion Patterns. *Pattern Recognition, Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on*, 2, 1994.

[50] J. Han and M. Kamber. *Data Mining: Concepts and Techniques.* Morgan Kaufmann, 2000.

[51] I. Haritaoglu, D. Harwood, and L. S. David. W4: Real-Time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):809–830, 2000.

[52] N. Herodotou, K. N. Plataniotis, and A. N. Venetsanopoulos. A Color Segmentation Scheme for Object-Based Video Coding. *Advances in Digital Filtering and Signal Processing, 1998 IEEE Symposium on*, pages 25–29, 1998.

[53] J. E. Herrero-Jaraba, C. Orrite-Uruñuela, D. Buldain, and A. Roy-Yarza. Human Recognition by Gait Analysis Using Neural Networks. *Proceedings of the International Conference on Artificial Neural Networks*, 2415:346–369, 2002.

[54] P. V. C. Hough. Method and Means for Recognizing Complex Patterns, December 18 1962. US Patent 3,069,654.

[55] W. Hu, T. Tan, L. Wang, and S. Maybank. A Survey on Visual Surveillance of Object Motion and Behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(3):334–352, 2004.

[56] P. S. Huang, C. J. Harris, and M. S. Nixon. Human Gait Recognition in Canonical Space using Temporal Templates. *In IEE Proceedings of Vision, Image and Signal Processing*, 146(2):93–100, 1999.

[57] M. C. Isard and A. C. Blake. CONDENSATION: Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[58] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima. Real-Time Estimation of Human Body Posture from Monocular Thermal Images. *In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 15–20, 1997.

[59] A. K. Jain, R. Bolle, and S. Pankanti. *Biometrics: Personal Identification in Networked Society.* Springer, 1999.

[60] A. K. Jain and B. Chandrasekaran. Dimensionality and Sample Size Considerations in Pattern Recognition Practice. *Handbook of Statistics*, 2:835–855, 1982.

[61] A. K. Jain and D. Zongker. Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[62] O. Javed and M. Shah. Tracking and Object Classification for Automated Surveillance. *In Proceedings of the European Conference on Computer Vision*, 4:343–357, 2002.

[63] G. Johansson. Visual Perception of Biological Motion and a Model for its Analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[64] A. Y. Johnson and A. F. Bobick. A Multi-View Method for Gait Recognition using Static Body Parameters. *In Proceedings of the 3rd International Conference on Audio-and Video-Based Biometric Person Authentication*, 716:301–311, 2001.

[65] S. J. Julier and J. K. Uhlmann. A New Extension of the Kalman Filter to Nonlinear Systems. *International Symposium Aerospace/Defense Sensing, Simulation and Controls*, 3, 1997.

[66] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger. Gait-Based Recognition of Humans using Continuous HMMs. *In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 321–326, 2002.

[67] I. A. Karaulova, P. M. Hall, and A. D. Marshall. A Hierarchical Model of Dynamics for Tracking People with a Single Video Camera. *In Proceedings of the 11th British Machine Vision Conference*, 1:352–361, 2000.

[68] R. F. Kleissen, J. H. Buurke, J. Harlaar, and G. Zilvold. Electromyography in the Biomechanical Analysis of Human Movement and its Clinical Application. *Gait Posture*, 8(2):143–158, 1998.

[69] J. Knapik, E. Harman, and K. Reynolds. Load Carriage using Packs: A Review of Physiological, Biomechanical and Medical Aspects. *Applied Ergonomics*, 27(3):207–216, 1996.

[70] L. T. Kozlowski and J. E. Cutting. Recognizing the Gender of Walkers from Point-Lights Mounted on Ankles: Some Second Thoughts. *Perception & Psychophysics*, 23(5):459, 1978.

[71] F. P. Kuhl and C. R. Giardina. Elliptic Fourier Features of a Closed Contour. *Computer Graphics and Image Processing*, 18(3):236–258, 1982.

[72] B. C. Kuo and D. A. Landgrebe. Nonparametric Weighted Feature Extraction for Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42(5):1096–1105, 2004.

[73] S. Kurakake and R. Nevatia. Description and Tracking of Moving Articulated Objects. *In Proceedings of the 11th IAPR International Conference on Pattern Recognition, Conference A: Computer Vision and Applications*, 1:491–495, 1992.

[74] J. L. Landabaso, L. Q. Xu, and M. Pardas. Robust tracking and object classification towards automated video surveillance. *In Proceedings of Internacional Conference on Image Analysis and Recognition*, 3212:463–470, 2004.

[75] V. F. Leavers. Which Hough Transform? *CVGIP: Image Understanding*, 58(2):250–264, 1993.

[76] L. Lee and WEL Grimson. Gait Analysis for Recognition and Classification. *In Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition.*, pages 148–155, 2002.

[77] A. J. Lipton, H. Fujiyoshi, and R. S. Patil. Moving Target Classification and Tracking from Real-Time Video. *In Proceedings of IEEE Workshop on Applications of Computer Vision*, 14, 1998.

[78] J. Little and J. Boyd. Recognizing People by their Gait: the Shape of Motion. *Videre: Journal of Computer Vision Research*, 1(2):2–32, 1998.

[79] N. Lynnerup and J. Vedel. Person Identification by Gait Analysis and Photogrammetry. *Journal of Forensic Sciences*, 50(1):112–118, 2005.

[80] D. Meyer, J. Denzler, and H. Niemann. Model Based Extraction of Articulated Objects in Image Sequences for Gait Analysis. *In Proceedings of the International Conference on Image Processing*, 3, 1997.

[81] T. B. Moeslund and E. Granum. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.

[82] T. B. Moeslund, A. Hilton, and V. Krüger. A Survey of Advances in Vision-Based Human Motion Capture and Analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, 2006.

[83] S. D. Mowbray and M. S. Nixon. Automatic Gait Recognition via Fourier Descriptors of Deformable Objects. *In Proceedings of the 4th Conference on Audio- and Video-Based Biometric Person Authentication*, pages 566–573, 2003.

[84] M. P. Murray. Gait as a Total Pattern of Movement. *American Journal of Physical Medicine*, 46(1):290–333, 1967.

[85] M. P. Murray, A. B. Drought, and R. C. Kory. Walking Patterns of Normal Men. *The Journal of Bone and Joint Surgery*, 46(2):335, 1964.

[86] E. Muybridge. *Muybridge's Complete Human and Animal Locomotion: All 781 Plates from the 1887 Animal Locomotion.* Courier Dover Publications, 1979.

[87] H. Ning, T. Tan, L. Wang, and W. Hu. People Tracking Based on Motion Model and Motion Constraints with Automatic Initialization. *Pattern Recognition*, 37(7):1423–1440, 2004.

[88] M. Nixon and A.S. Aguado. *Feature Extraction & Image Processing.* Academic Press, 2007.

[89] M. S. Nixon and J. N. Carter. Advances in Automatic Gait Recognition. *In Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 139–144, 2004.

[90] M. S. Nixon and J. N. Carter. Automatic Recognition by Gait. *In Proceedings of the IEEE*, 94(11):2013–2024, 2006.

[91] M. S. Nixon, J. N. Carter, J. M. Nash, P. S. Huang, D. Cunado, and S. V. Stevenage. Automatic Gait Recognition. *Motion Analysis and Tracking (Ref. No. 1999/103), IEE Colloquium on*, page 3, 1999.

[92] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait (The Kluwer International Series on Biometrics)*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2005.

[93] S. A. Niyogi and E. H. Adelson. Analyzing and Recognizing Walking Figures in XYT. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[94] M. A. Nurse, M. Hulliger, J. M. Wakeling, B. M. Nigg, and D. J. Stefanyshyn. Changing the Texture of Footwear can Alter Gait Patterns. *Journal of Electromyography & Kinesiology*, 15(5):496–506, 2005.

[95] D. D. Pascoe, D. E. Pascoe, Y. T. Wang, D.M. Shim, and C. K. Kim. Influence of Carrying Book Bags on Gait Cycle and Posture of Youths. *Ergonomics*, 40(6):631–640, 1997.

[96] F. J. Perales. Human Motion Analysis and Synthesis using Computer Vision and Graphics Techniques. State of Art and Applications. *In Proceedings of the world multiconference on systemics, cybernetics and informatics (SCI2001)*, 2001.

[97] S. Pheasant. *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. Taylor & Francis, 1996.

[98] P. J. Phillips. Human Identification Technical Challenges. *In Proceedings of the International Conference on Image Processing*, 1, 2002.

[99] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[100] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. The gait Identification Challenge Problem: Data Sets and Baseline Algorithm. *In Proceedings of 16th International Conference on Pattern Recognition*, 1:385–388, 2002.

[101] P. Pudil, J. Novovičová, and J. Kittler. Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994.

[102] A. Punakallio, S. Lusa, and R. Luukkonen. Protective Equipment Affects Balance Abilities Differently in Younger and Older Firefighters. *Aviation, Space, and Environmental Medicine*, 74(11):1151–1156, 2003.

[103] S. Rahmatalla, H. Kim, M. Shanahan, and C. C. Swan. Effect of Restrictive Clothing on Balance and Gait using Motion Capture and Dynamic Analysis. *SAE Transactions Journal of Passenger Cars: Electronic and Electrical Systems*, 114:713–722, 2005.

[104] K. Rohr. Towards Model-Based Recognition of Human Movements in Image Sequences. *CVGIP. Image understanding*, 59(1):94–115, 1994.

[105] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanID Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[106] A. Shio and J. Sklansky. Segmentation of People in Motion. *In Proceedings of the IEEE Workshop on Visual Motion*, pages 325–332, 1991.

[107] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter. On a Large Sequence-Based Human Gait Database. *In Proceedings of Recent Advances in Soft Computing*, pages 66–71, 2002.

[108] L. Sloman, M. Berridge, S. Homatidis, D. Hunter, and T. Duck. Gait Patterns of Depressed Patients and Normal Subjects. *American Journal of Psychiatry*, 139(1):94–7, 1982.

[109] P. Somol, P. Pudil, J. Novovičová, and P. Paclík. Adaptive Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 20(11-13):1157–1163, 1999.

[110] N. Spencer and J. N. Carter. Towards Pose Invariant Gait Reconstruction. *In Proceedings of IEEE International Conference on Image Processing*, 3, 2005.

[111] C. Stauffer and W. E. L. Grimson. Learning Patterns of Activity using Real-Time Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

[112] S. V. Stevenage, M. S. Nixon, and K. Vince. Visual Analysis of Gait as a Cue to Identity. *Applied Cognitive Psychology*, 13(6):513–526, 1999.

[113] A. Sundaresan, A. RoyChowdhury, and R. Chellappa. A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences. *In Proceedings of the International Conference on Image Processing*, 2:93–96, 2003.

[114] R. Tanawongsuwan and A. Bobick. Modelling the Effects of Walking Speed on Appearance-Based Gait Recognition. *in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2.

[115] R. Tanawongsuwan and A. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 726–731, 2001.

[116] R. Tanawongsuwan and A. Bobick. Performance Analysis of Time-Distance Gait Parameters under Different Speeds. *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 715–724, 2003.

[117] N. F. Troje, C. Westhoff, and M. Lavrov. Person Identification from Biological Motion: Effects of Structural and Kinematic Cues. *Perception & Psychophysics*, 67(4):667–675, 2005.

[118] I. R. Vega and S. Sarkar. Statistical Motion Model Based on the Change of Feature Relationships: Human Gait-Based Recognition. 2003.

[119] G. V. Veres, M. S. Nixon, and J. N. Carter. Modelling the Time-Variant Covariates for Gait Recognition. *in Proceedings of 5th International conference on Audio-and Video-Based Viometric Person Authentication*, pages 597–606, 2005.

[120] G. V. Veres, M. S. Nixon, L. Middleton, and J. N. Carter. Fusion of Dynamic and Static Features for Gait Recognition over Time. *In Proceedings of 7th International Conference on Information Fusion*, 2, 2005.

[121] D. K. Wagg. *Local and Global Models for Articulated Motion Analysis*. PhD thesis, University of Southampton, 2005.

[122] D. K. Wagg and M. S. Nixon. On Automated Model-Based Extraction and Analysis of Gait. *In Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 11–16, 2004.

[123] L. Wang, W. Hu, and T. Tan. A New Attempt to Gait-Based Human Identification. *In Proceedings of the International Conference on Pattern Recognition*, 1:115–118, 2002.

[124] L. Wang, W. Hu, and T. Tan. Recent Developments in Human Motion Analysis. *Pattern Recognition*, 36(3):585–601, 2003.

[125] L. Wang, H. Ning, T. Tan, and W. Hu. Fusion of Static and Dynamic Body Biometrics for Gait Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):149–158, 2004.

[126] L. Wang, T. Tan, W. Hu, and H. Ning. Automatic Gait Recognition Based on Statistical Shape Analysis. *IEEE Transactions on Image Processing*, 12(9):1120–1131, 2003.

[127] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

[128] Y. T. Wang, D. D. Pascoe, and W. Weimar. Evaluation of Book Backpack Load during Walking. *Ergonomics*, 44(9):858–869, 2001.

[129] A. R. Webb. *Statistical Pattern Recognition*. Hodder Arnold, 1999.

[130] G. Welch and G. Bishop. An Introduction to the Kalman Filter. *ACM SIGGRAPH 2001 Course Notes*, 2001.

[131] M. Whittle. *Gait Analysis: An Introduction*. Butterworth-Heinemann, 2002.

[132] D. A. Winter. *The Biomechanics and Motor Control of Human Movement*. John Wiley & Sons, second edition, 1990.

[133] D. A. Winter. *The Biomechanics and Motor Control of Human Gait: Normal Elderly and Pathological*. University of Waterloo Press, 1991.

[134] D. A. Winter. Human Balance and Posture Control during Standing and Walking. *Gait and Posture*, 3(4):193–214, 1995.

[135] M. H. Woollacott and P. F. Tang. Balance Control during Walking in the Older Adult: Research and its Implications. *Physical Therapy*, 77(6):646, 1997.

[136] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[137] C. Y. Yam, M. S. Nixon, and J. N. Carter. On the Relationship of Human Walking and Running: Automatic Person Identification by Gait. *in Proceedings of the International Conference on Pattern Recognition*, 2002.

[138] C. Y. Yam, M. S. Nixon, and J. N. Carter. Automated Person Recognition by Walking and Running via Model-Based Approaches. *Pattern Recognition*, 37(5):1057–1072, 2004.

[139] J. H. Yoo, M. S. Nixon, and C. J. Harris. Extraction and Description of Moving Human Body by Periodic Motion Analysis. *In Proceedings of ISCA 17th International Conference on Computers and Their Applications*, pages 110–113, 2002.

[140] D. Zhang and G. Lu. Evaluation of Similarity Measurement for Image Retrieval. *In Proceedings of the International Conference on Neural Networks and Signal Processing*, 2, 2003.

[141] D. Zongker and A. Jain. Algorithms for Feature Selection: An evaluation. *Proceedings of the 13th International Conference on Pattern Recognition*, 2, 1996.