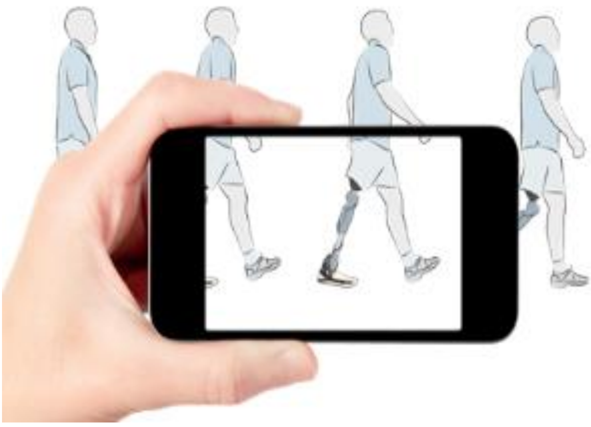
UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO



Gait Analysis in Unconstrained Environments

Tanmay Tulsidas Verlekar

Supervisor: Doctor Paulo Luís Serras Lobato Correia

Co-supervisor: Doctor Luís Eduardo de Pinho Ducla Soares

Thesis approved in public session to obtain the PhD Degree in
Electrical and Computer Engineering

Jury final classification: Pass with Distinction and Honour

2019

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Gait Analysis in Unconstrained Environments

Tanmay Tulsidas Verlekar

Supervisor: Doctor Paulo Luís Serras Lobato Correia
Co-supervisor: Doctor Luís Eduardo de Pinho Ducla Soares

Thesis approved in public session to obtain the PhD Degree in
Electrical and Computer Engineering

Jury final classification: Pass with Distinction and Honour

Jury

Chairperson: Doctor Mário Alexandre Teles de Figueiredo, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Hugo Pedro Martins Carriço Proença, Faculdade de Engenharia, Universidade da Beira Interior

Doctor Ana Luísa Nobre Fred, Instituto Superior Técnico, Universidade de Lisboa

Doctor Alexandre José Malheiro Bernardino, Instituto Superior Técnico, Universidade de Lisboa

Doctor Paulo Luís Serras Lobato Correia, Instituto Superior Técnico, Universidade de Lisboa

Doctor Hugo Humberto Plácido da Silva, Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal

Funding Institutions
This research has been made possible with funding from the Fundação para a Ciência e a Tecnologia, Instituto de Telecomunicações, Instituto Superior Técnico.

2019

ABSTRACT

Gait can be defined as the individuals' manner of walking. Its analysis can provide significant information about their identity and health, opening a wide range of possibilities in the field of biometric recognition and medical diagnosis. In the field of biometric, the use of gait to perform recognition can provide advantages, such as acquisition from a distance and without the cooperation of the individual being observed. In the field of medicine, gait analysis can be used to detect or assess the development of different gait related pathologies. It can also be used to assess neurological or systemic disorders as their effects are reflected in the individuals' gait.

This Thesis focuses on performing gait analysis in unconstrained environments, using a single 2D camera. This can be a challenging task due to the lack of depth information and self-occlusions in a 2D video sequence. The Thesis explores the use of gait, to perform biometric recognition and pathology detection and classification by reviewing the state-of-the-art and presenting novel taxonomies to organise the systems.

In the field of biometrics, the work done in this Thesis improves the performance of the recognition systems by proposing two novel gait representations. It also addresses the problems faced by recognition systems in unconstrained environments, such as change in the viewpoint of the camera and change in the appearance of the individuals being observed, presenting three novel systems to detect the viewpoint of the camera and a system to tackle appearance change. Finally, the Thesis explores the possibility of obtaining gait features from the shadow cast by the individuals, presenting two systems to rectify the distortion and deformation in the shadow silhouettes and a system to detect if the shadow is usable. It also presents two datasets to evaluate these systems.

In the field of medicine, this Thesis presents a novel system to obtain biomechanical features, from a video sequence captured with a 2D camera, with a high level of accuracy, while also being robust to viewpoint change. To evaluate the system the Thesis presents a dataset containing sequences acquired from a 2D camera and the "gold standard" motion capture system. The Thesis also explores the ability of gait to classify different gait related pathologies. It presents two novel systems that perform classification of gait across different gait related pathologies using biomechanical features and deep convolutional neural networks.

A comprehensive evaluation of the proposed systems and comparison with the state-of-the-art highlight the advantages of the proposed systems for biometric recognition and pathology classification.

Keywords: *Gait analysis, Biometric recognition, Shadow analysis, Biomedical analysis, Pathology classification*

RESUMO

A marcha pode ser definida como a maneira de andar dos indivíduos. A análise automática da marcha pode fornecer informação sobre a identidade e a saúde de um individuo, podendo ser explorada no âmbito do reconhecimento biométrico e do diagnóstico médico. No campo da biometria, o reconhecimento usando características da marcha pode trazer vantagens, como a aquisição à distância e sem a cooperação explícita do indivíduo. No campo da medicina, a análise da marcha pode ajudar a detetar e avaliar o desenvolvimento de diferentes patologias. Também pode ser usado para avaliar perturbações de origem neurológica ou sistémica que se refletem na marcha dos indivíduos.

Esta tese foca-se na análise da marcha em ambientes sem grandes restrições, usando uma única camara de vídeo 2D. Esta análise apresenta alguns desafios, pois numa sequência de vídeo 2D não está disponível informação de profundidade ou de auto-oclusões. A tese analisa as características da marcha para realizar reconhecimento biométrico e para a deteção e classificação de patologias da marcha, incluindo uma revisão do estado da arte e apresentando novas taxonomias para organizar as soluções existentes.

No campo da biometria, o trabalho desenvolvido nesta tese contribui para melhorar o desempenho dos sistemas de reconhecimento, propondo duas novas representações da marcha. Também aborda os problemas enfrentados pelos sistemas de reconhecimento em ambientes sem restrições, como a mudança no ponto de vista da camara e as alterações na aparência dos indivíduos observados, apresentando três novos sistemas para detetar o ponto de vista da camara e um sistema para lidar com as alterações de aparência. Por fim, a tese explora a possibilidade de obter características da marcha a partir da sombra projetada pelos indivíduos, apresentando dois sistemas para retificar a distorção e deformação nas silhuetas de sombra e um sistema para reconhecimento baseado na sombra. Também apresenta duas bases de dados para avaliar estes sistemas.

No campo da medicina, esta tese apresenta um novo sistema para obtenção de características biomecânicas a partir de uma sequência de vídeo, capturada com uma camara 2D, com alto nível de precisão, além de robusta à mudança de ponto de vista. Para avaliar o sistema, a tese apresenta uma base de dados contendo sequencias adquiridas em simultâneo por uma camara 2D e por um sistema de captura de movimento 3D, considerado como a referência para comparações. A tese também explora a marcha para classificar diferentes patologias. Apresenta dois novos sistemas que detetam e classificação diferentes patologias da marcha, utilizando características biomecânicas e redes neuronais convolucionais profundas.

É apresentada uma avaliação abrangente dos sistemas propostos, em comparação com o estado-da-arte, sendo de destacar as vantagens dos sistemas propostos para reconhecimento biométrico e para deteção e classificação de patologias da marcha.

Palavras-chave: Análise da marcha, Reconhecimento biométrico, Análise de sombras, Análise biomédica, Classificação de patologias

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my advisors Prof. Paulo Luís Serras Lobato Correia and Prof. Luís Eduardo de Pinho Ducla Soares for their continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this Thesis. I could not have imagined having better advisors and mentors for my Ph.D study.

My sincere thanks also go to Dr. Kurt Claeys, Dr. Hans Hallez and Dr. Chang-Tsun Li who provided me opportunities to join their groups and access their laboratories and research facilities.

This work would not have been possible without the financial support of Instituto de Telecomunicações (IT), Instituto Superior Técnico (IST) and Fundação para a Ciência e a Tecnologia (FCT). In Addition, I am grateful for the facilities provided by IT and for the assistance offered by the IT staff.

I would like to thank my colleagues in IT and friends for stimulating discussions and providing moral support.

Last but not the least; I would like to thank my family: my parents and my brother for supporting me spiritually throughout writing this thesis and my life in general.

Tanmay Tulsidas Verlekar

Lisbon, February 2019

TABLE OF CONTENTS

Abstract v
Resumo vii
Acknowledgement ix
List of figures xv
List of tables xix
List of Abbreviations xxi
Part I: Gait analysis basics and the state-of-the-art 1
1 Introduction 3
1.1 Context and motivation 3
1.1.1 Gait-based biometric recognition 4
1.1.2 Gait-based pathology classification 6
1.2 Objectives 7
1.3 Contributions 8
1.4 Thesis structure 11
2 Gait analysis: basic concepts and acquisition systems 13
2.1 Introduction 13
2.2 Gait structure and representation 13
2.3 Gait acquisition systems 16
2.3.1 Systems with wearable sensors 16
2.3.2 Systems with non-wearable sensors 17
2.4 Challenges to 2D vision-based gait analysis in unconstrained environments 19
2.4.1 Individual-related factors 19
2.4.2 Camera-related factors 20
2.4.3 Light source-related factors 21
2.4.4 Environment-related factors 22
2.4.5 Combination of factors 22
2.5 System architecture and evaluation metrics 23
3 Gait analysis: state-of-the-art 27
3.1 Introduction 27
3.2 Gait-based biometric recognition systems 27
3.2.1 Model-based systems 27
3.2.2 Appearance-based systems 29

3.2.3 Discussion 33
3.2.4 Datasets 33
3.3 Gait-based pathology classification systems 35
3.3.1 Model-based systems 36
3.3.2 Appearance-based systems 36
3.3.3 Discussion 37
3.3.4 Datasets 37
Part II: Gait-based biometric recognition 39
4 Gait representation proposals 41
4.1 Introduction 41
4.2 Gait dissimilarity vector 41
4.3 Sparse error gait image 43
4.4 Performance evaluation 45
5 Proposals to tackle changes in the camera viewpoint 49
5.1 Introduction 49
5.2 Viewpoint detection using a perceptual hash 49
5.3 Viewpoint detection using gait texture image contours 50
5.4 Viewpoint detection in the presence of shadow 52
5.5 Performance evaluation 54
6 Proposal to tackle change in the appearance of the individuals 59
6.1 Introduction 59
6.2 Tackling appearance changes using GEI decomposition 59
6.3 Performance evaluation 61
7 Proposals for rectification of shadow silhouettes 65
7.1 Introduction 65
7.2 Rectification of silhouettes using robust principal component 65
7.3 Rectification of silhouettes using 4-point correspondence 67
7.4 Shadow type identification 69
7.5 The IST shadow datasets 71
7.5.1 The IST shadow gait dataset 71
7.5.2 The IST shadow type dataset 72
7.6 Performance evaluation 72
Part III: Gait-based pathology classification 77
8 Proposal for accurate estimation of temporal biomechanical gait features 79
8.1 Introduction 79
8.2 Estimation of stance time, swing time and gait cycle time using IC and TO 79

8.3 IST-KUL dataset acquisition ... 82
8.4 Performance evaluation ... 83
9 Proposal for classification of gait related pathologies ... 87
9.1 Introduction ... 87
9.2 Classification of pathologies using “hand-crafted” biomechanical features ... 87
9.3 Classification of pathologies using a deep CNN and transfer learning ... 93
9.4 Performance evaluation ... 95
9.4.1 Performance evaluation of the biomechanical feature-based system ... 95
9.4.2 Performance evaluation of the deep CNN-based system ... 98
Part IV: Conclusion ... 103
10 Final remarks ... 105
10.1 Introduction ... 105
10.2 Summary of contributions ... 105
10.2.1 Gait-based biometric recognition ... 105
10.2.2 Gait-based pathology classification ... 107
10.3 Future research directions ... 108
10.3.1 Gait-based biometric recognition ... 108
10.3.2 Gait-based pathology classification ... 109
References ... 113

LIST OF FIGURES

Figure 1.1: Application of gait analysis in (a) medical diagnosis, (b) biometric recognition and (c) sports [7]. 3
Figure 1.2: Biometric recognition systems performing identification and verification [20]. 5
Figure 1.3: Illustrations of (a) a gait representation (b) change in viewpoint, (c) occlusion caused by change in appearance, (d) shadow cast by an individual [23]. 6
Figure 1.4: Illustration of (a) spastic gait, (b) scissors gait, (c) steppage gait, (d) waddling gait, (e) propulsive gait [25]. 7
Figure 2.1: Key events and phases that occur during a gait cycle [39]. 14
Figure 2.2: Illustration of spatial features, including (a) distance and (b) angular features. 15
Figure 2.3: Proposed taxonomy for the state-of-the-art gait analysis systems. 16
Figure 2.4: Examples of wearable sensors: (a) force sensitive resistors [43]*, (b) inertial measurement units* [44]*, (c) electromyograph* [45]. 17
Figure 2.5: Non-wearable sensor-based systems: (a) floor-based system [49], (b) marker-based vision system [50], (c) marker-less vision system [51]. 18
Figure 2.6: Steps in 2D vision-based gait analysis (a) background subtraction, (b) feature estimation. 19
Figure 2.7: Factors affecting gait analysis using 2D vision-based systems. 20
Figure 2.8: Examples of silhouettes for: (a) normal gait, (b) gait partly occluded by a bag, (c) gait altered by speed and (d) gait affected by health. 21
Figure 2.9: Examples of shadow silhouette deformations caused due to field of view, skewness and scale changes. 21
Figure 2.10: Examples of (a) sharp and (b) diffused shadows cast by the walking individual. 22
Figure 2.11: Different viewpoints of the same individual: (a) frontal view, (b) side view, (c) view at the start of the sequence, (d) view at the end. 23
Figure 2.12: Change in the shadow cast by the same individual: (a, b) when the individual's position w.r.t. the light source changes, and at the (b) start and (c) end of a linear walking sequence. 23
Figure 2.13: Architecture of the 2D vision-based systems for biometric recognition and pathology classification. 24
Figure 2.14: Illustration of a confusion matrix 24
Figure 2.15: Illustration of a (a) ROC and (b) CMC curve. 25
Figure 3.1: Proposed taxonomy for the state-of-the-art vision-based systems. 28
Figure 3.2: Gait representations obtained from (a) 3D model-based system, (b) feature invariant systems [75]. 29
Figure 3.3: Illustration of a GEI and GEnI. 30
Figure 3.4: GEIs of an individual corresponding to different angles representing the viewpoints of the camera. 31
Figure 3.5: GEIs of an individual altered by wearing a coat and carrying a bag. 31
Figure 3.6: Image captured using an overhead camera in an unconstrained environment. 32
Figure 3.7: CASIA B dataset representing (a) change in camera's view point and (b) appearance change, (c) OU-ISIR LP dataset representing viewpoint change, (d) KY IR shadow dataset. 35
Figure 3.8: Gait representing (a) Hemiplegia, (b) Parkinson's diseases, (c) Cerebellar ataxia, (d) Foot drop and (e) Sensory ataxia [111]. 35

Figure 3.9: Examples of GEI belonging to (a) DAI 2 and (b) INIT dataset. 38
Figure 4.1: Proposed gait dissimilarity vector representation. 42
Figure 4.2: Gait recognition using the proposed gait dissimilarity vector (GDV) representation. 42
Figure 4.3: (a) Obtaining SEGI by applying RPCA to the GEIs, (b) SEGI representation and their intensity value plot when considering GEIs from the same individual and (c) different individuals. 43
Figure 4.4: Proposed system architecture (SEGI). 44
Figure 4.5: Change in recognition accuracy w.r.t. increase in the number of registered sequences per individual. 47
Figure 5.1: Output of the intermediate steps of the PHash function (a) GEI, (b) selected leg region, (c) DCT coefficients, (d) matrix representing the hash bits. 50
Figure 5.2: Architecture of the proposed viewpoint detection system based on PHash. 50
Figure 5.3: Architecture of the proposed viewpoint detection system based on GTI contour. . 51
Figure 5.4: Construction of a GTI 51
Figure 5.5: Example of (a) selecting the feet position from the GTI contour and (b) application of PCA to the selected feet positions 52
Figure 5.6: GTI representing (a) 126˚ viewpoint, (c) 0˚ viewpoint, and the corresponding vertical sums of intensities for (b) 126˚ viewpoint and (d) 0˚ viewpoint GTI. 52
Figure 5.7: Illustration of (a) success and (b) failure of the system presented in [100] 53
Figure 5.8: Fitting a line through the identified feet positions using RANSAC. 53
Figure 5.9: Illustration of the intermediate steps of the proposed system depicting (a) feet position detection, (b) line fitting using RANSAC. 54
Figure 6.1: Architecture of the proposed gait recognition system robust to appearance changes, based on GEI decomposition 59
Figure 6.2: Illustration of the operation of the gait recognition system robust to appearance changes based on GEI decomposition: (a) average GEI image, (b) example of an unaltered GEI, (c) example of a GEI altered by wearing a coat, (d) binary mask representing the altered parts of the GEI, (e) illustration of the GEI sections considered, (f) GEI sections selected for recognition. 60
Figure 6.3: Performance of the proposed system with respect to the number of sections N. .. 61
Figure 6.4: Lateral and frontal examples of GEIs obtained from (a, b) normal, (c, d) coat and (e, f) bag sequences. 62
Figure 7.1: Change in viewpoint of the camera caused by observing individuals walking long distances. 65
Figure 7.2: Proposed Shadow Silhouette rectification system: (a) flipped shadow silhouette, (b) GTI_s, (c) $GTIso$, (d) rectified shadow silhouette. 67
Figure 7.3: (a) mid-stance/mid-swing detection, (b) head and feet positions selection using the proposed system. 68
Figure 7.4: The head and feet positions in the observed view (red) and their corresponding mapping in the canonical view (blue). 68
Figure 7.5: Illustration of (a) silhouette, (b) background and (c) foreground image. 70
Figure 7.6: Intensity ratio histogram plot and the corresponding images for (a, b) sharp and (c, d) diffused shadows 71
Figure 7.7: Illustration of the two different type of plots obtained for (a) sharp, (b) diffused shadows. 71
Figure 7.8: IST shadow gait dataset acquisition setup: (a) camera position, and (b) walking directions. 72

Figure 7.9: IST shadow type dataset acquisition setup (a) camera position (b) captured types of shadow. 72
Figure 7.10: Features used by the state-of-the-art and proposed systems (a) shadow silhouettes; (b) GEI, GEI obtained using rectified silhouettes using (c) TILT and (d) 4-point correspondence system. 74
Figure 7.11: Silhouettes at the start, middle and end of a gait sequence. (a, b, c) represent body silhouettes; (d, e, f) represent shadow silhouettes and (g, h, i) rectified shadow silhouettes. ... 76
Figure 8.1: Architecture of the proposed temporal features estimation system. 80
Figure 8.2: Illustration of the proposed foot flat identification steps: (a) GTI, (b) selected feet region of the GTI, (c) thresholding, (d) resulting “foot flat image” 80
Figure 8.3: Selection of candidate frames for the estimation of initial contact and toe off. 81
Figure 8.4: Examples of intermediate steps in IC and TO estimation: (a), (b) selection of foot from CIC and CTO, (c) flat feet selection, (d) overlap between flat foot and CIC/CTO feet, estimation of: (e) Initial contact (f) toe off. 82
Figure 8.5: Camera setup for the IST-KUL dataset. 83
Figure 8.6: Percentage of correctly estimated (a) ICs and (b) TOs w.r.t. the distance from the ground truth. 84
Figure 9.1: Architecture of the proposed gait-based pathology classification system, using “hand-crafted” biomechanical features. 88
Figure 9.2: Silhouettes belonging to a healthy individual and (b) an individual suffering from Parkinson’s diseases. 88
Figure 9.3: Plot representing the distance between feet for a gait cycle. 88
Figure 9.4: Intermediate steps for step length estimation: (a) segmented feet silhouettes between two initial contacts; (b) AFI obtained by averaging the feet silhouettes, (c) position of the foot flat obtained by applying a threshold, (d) centroids of foot flats obtained for the entire video sequence. 89
Figure 9.5: Plot representing the foot flat overlap ratio (top) and the corresponding foot and walking individual’s silhouettes (bottom). 91
Figure 9.6: Half cycle GEI computed using (a) impaired and (c) healthy gait silhouettes, and (b, d) the corresponding GEnI representations, where entropy highlights the amount of movement. 92
Figure 9.7: GEI highlighting shift in COG (middle silhouette point in red) with respect to the COS (lower silhouette point in blue) and the orientation of the torso. 93
Figure 9.8: Feature extraction (VGG-19) architecture. 94
Figure 10.1: Illustration of an individual (a) walking along a straight line, (b) walking along a curved trajectory 109
Figure 10.2: Shadow cast throughout the day. 109
Figure 10.3: Output of the 3D pose system [142]. 110
Figure 10.4: Deterioration of gait due to Parkinson’s diseases. 110

LIST OF TABLES

Table 3.1: Summary of available datasets for biometric gait recognition 34
Table 3.2: Summary of available datasets for pathology classification 37
Table 4.1: Recognition accuracy of the proposed and state-of-the-art gait representations. ... 47
Table 5.1: Correct viewpoint detection accuracy (%) for N - Normal, C - Coat and B - Bag sequences.. 55
Table 5.2: Confusion matrix for the proposed PHash system operating on OU-ISIR LP dataset.56
Table 5.3 Recognition accuracy of recognition system with and without viewpoint detection. 57
Table 6.1: Recognition accuracy (%) of the proposed and the state-of-the-art systems robust to appearance change. 62
Table 6.2: Recognition accuracy (%) of the proposed system w.r.t. different viewpoints of the camera... 63
Table 7.1: Recognition accuracy (%) of the state-of-the-art and proposed systems that rely on shadow silhouettes. 73
Table 7.2: Recognition accuracy (%) of the system using GEIs composed of body silhouettes. .. 74
Table 7.3: Recognition accuracy (%) of the system using GEIs composed of shadow silhouettes. 74
Table 7.4: Recognition accuracy (%) of the system using GEIs composed of rectified shadow silhouettes obtained using the proposed 4-point correspondence system. 75
Table 8.1: IC and TO estimation with an error margin of ±2 frames w.r.t. the ground truth. 84
Table 8.2: IC and TO estimation with an error margin of ±1 frame across different viewpoints. 85
Table 8.3: Comparison between the proposed and motion capture system. 86
Table 9.1: Two sample t-test with unequal variances and significance level of 0.05 performed between normal and impaired gait... 96
Table 9.2: Mean and standard deviation of all the observed gait features belonging to different gait related pathologies. 97
Table 9.3: Classification accuracy (%) of the proposed “hand-crafted” biomechanical feature-based system and state-of-the-art systems in classifying between FB, RL, LL and NM gait....... 98
Table 9.4: Confusion matrix (%) for the proposed biomechanical features-based system. 98
Table 9.5: VGG-19 configurations and the corresponding classification accuracy (%). 100
Table 9.6: Classification accuracy (%) of the system proposed and the state-of-the-art. 101
Table 9.7: Confusion matrix (%) for the system proposed in Section 9.2 operating on DAI 2 dataset... 101

LIST OF ABBREVIATIONS

2D - 2 Dimensional

3D - 3 Dimensional

AFI - Average feet image

AOM - Amount of movement

CGI - Chrono gait image

CIC - Candidate for initial contact

CMC - Cumulative match characteristic

CNN - Convolutional neural network

COG - Centre of gravity

CONV - convolutional

COS - Centre of support

CTO - Candidate for toe off

DCT - Discrete cosine transform

FDEI - Frame difference energy image

FC - fully connected

FCT - Fundação para a Ciência e a Tecnologia

FFR - Foot flat ratio

FN - False negative

FP - False positive

FPS - Frames per second

GDV - Gait dissimilarity vector

GEI - Gait energy image

GEnI - Gait entropy image

GTI - Gait texture image

IC - Initial contact

IMU - Inertial measurement unit

IST - Instituto Superior Técnico

IT - Instituto de Telecomunicações

K-NN - k-nearest neighbour

KUL - Katholieke Universiteit Leuven

LDA - Linear discriminant analysis

MGI - multiscale gait image

NSC - Normalised step count

OT - Orientation of the torso

PCA - Principal component analysis

PHash- Perceptual hash

RPCA - Robust Principal component analysis

ROC - Receiver operating characteristic

SEGI - Sparse error gait image

SL - Step length

SVM - Support vector machine

TN – True negative

TP - True positive

TO - Toe off

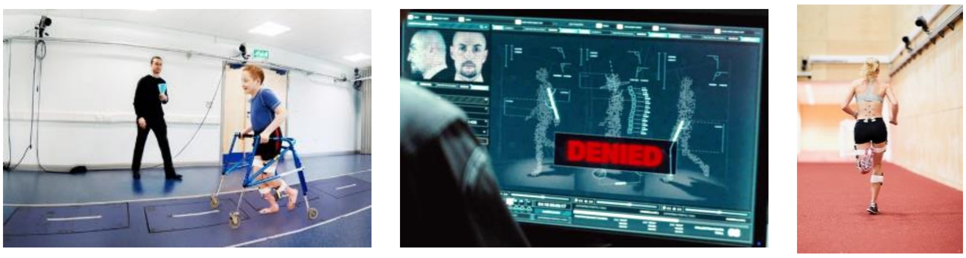
VTM - View transformation model

Part I: Gait analysis basics and the state-of-the-art

1 INTRODUCTION

1.1 CONTEXT AND MOTIVATION

Gait can be defined as a coordinated, cyclic combination of movements that results in the individuals' locomotion. Being a cognitive task, its analysis can provide useful information about the individuals' health, identity, gender or walking patterns [1]. The analysis of gait involves a systematic study of the individuals' motion by measuring body movements, activity of the muscles and other biomechanical aspects of the gait, to draw quantitative and interpretative conclusions from the resulting patterns. The information obtained has a wide range of application in the fields of medicine, biometrics, forensics and sports, as illustrated in Figure 1.1. In medicine, gait analysis can be used to detect or assess the development of different gait related pathologies, for instance resulting from neurological or systemic disorders, diseases, injuries or simply ageing [2]. Being a unique trait, gait can be used to recognise the identity of the individuals being observed, in biometric and forensic applications [3]. In sports, it can help athletes perform more efficiently by identifying posture-related or movement-related problems and can also help prevent injuries [4]. Gait analysis can also be used for determining an individuals' sex, gender or age group with a high level of accuracy [5] and can even reveal more complex information, such as body weight, feelings and emotions [6].



(a) (b) (c)

Figure 1.1: Application of gait analysis in (a) medical diagnosis, (b) biometric recognition and (c) sports [7].

Gait analysis has a rich history dating back to at least 350 BC, when Aristotle first kept written records of human and animal gait [8]. It was not until 1679 that Borelli associated mathematical models to the muscular movement and body dynamics of individuals [9]. Further studies about human gait were made by Wilhelm and Eduard Weber in 1836, describing the movements of the centre of gravity during walking and the pendulum-like behaviour of the legs during forward motion [10]. The progress made in the analysis of gait was slow as the observations were made through the naked eye. A significant improvement in the analysis of gait occurred after the invention of cameras, such as Marey's chronophotographic gun that could capture multiple frames per second (fps) [11]. This device, created in 1882, could capture video with a frame rate of 12 fps, providing a better insight into human gait, by capturing motion that is not perceptible to the naked eye. In 1895, Braune and Fischer improved on the study of the biomechanics of gait, by analysing the individuals' mass, volume and the centre of mass [12]. Braune and Fischer's study on gait was so complete that it was used as a reference for more than half a century.

The end of World War II led to a renewed interest in gait analysis, to assist injured soldiers and amputees returning from the war. This allowed Verne Inman and Howard Eberhart to study and present a unifying theory of gait following their observation that individuals move in such a way that the energy cost of moving the centre of gravity is the minimum [13]. In 1968, Jacqueline Perry identified the events that occur within a gait cycle and presented the division of the gait cycle into five stance and three swing phase events [14] - see Section 2.2. David Sutherland later improved on the model by substituting the three stance phase events with initial double support, single leg stance, and second double support [15].

In 1970, Patricia Murray became a world-renowned researcher for her contributions in gait analysis. She measured the kinematics of normal and impaired gait, recorded the change in the total movement pattern of normal individuals from infancy to old age and studied gait disturbances in individuals with neuromuscular and musculoskeletal pathologies. Her pioneering work in identifying features of normal gait that are consistent within an individual, but varied between individuals, provided the basis for forensic and biometric gait recognition [16]. The stability of normal gait and its distinctiveness between individuals was later demonstrated by Johansson, Cutting and Kozlowski in 1977, emphasizing its use in biometric recognition [17]. However, accurate localisation of key anatomical positions of the human body in 3D space was only possible after the development of the motion capture system in 1984 - see Section 2.3.2. Such systems are still considered as the gold standard in the acquisition of gait information [18].

Currently, gait analysis for medical diagnosis is performed in laboratories, under the supervision of clinical professionals. In addition, the biometric analysis of gait is often performed in constrained conditions, such as recognising the individuals walking along a predefined path. These limitations of the existing systems prevent their usage in a daily life setting or in generic surveillance scenarios. Thus, the need to perform automatic and objective gait analysis in unconstrained environments has gained significant interest of the research community. This Thesis deals with gait analysis in unconstrained environments, presenting systems that address several limitations of the current state-of-the-art. It explores the use of gait, acquired from a 2D camera, to perform the tasks of biometric recognition and pathology detection and classification. More details about these two tasks are given in the following sections.

1.1.1 Gait-based biometric recognition

Biometrics can be defined as the process of obtaining descriptive features based on either the behavioural, physical or physiological characteristics of the individuals' body, which can distinguish them uniquely among other individuals [19]. A biometric recognition system automatically compares the features obtained from the observed biometric traits, against a database containing features recorded during enrolment, to either verify the identity of the individuals being observed or to identify them – see Figure 1.2. To verify the claimed identity of an individual, a biometric recognition system performs a one-to-one match of the features against those features stored in the database, to either confirm or deny the claimed identity. On the other hand, identification is performed when a biometric recognition system assigns an identity to an individual from the identities registered in the database. If the system is forced to infer the identity from a list of individuals registered in the database, it is called a closed-set identification system. If the system can return a no match response, it is called an open-set identification system.

This Thesis explores biometric recognition systems because they have a significant advantage over traditional recognition systems, such as identity cards, passwords or pin numbers, as they cannot be forgotten and cannot be easily stolen. The novel systems presented in this Thesis are closed-set identification systems, as they have a wider range of application than verification systems.

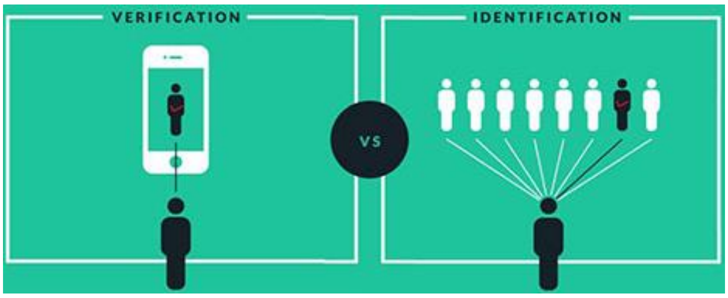


Figure 1.2: Biometric recognition systems performing identification and verification [20].

The features used by the biometric recognition systems can be obtained from physiological or behavioural traits. Physiological traits are related to the shape of the body and include face, ear and fingerprint. Behavioural traits are related to the pattern of behaviour of the individuals, such as typing rhythm, voice and signature. These traits are unique to the individuals, to a varying degree, but depending on their application can obtain excellent recognition results [21]. The applications include physical access control, logging attendance or personal identification. However, most biometric traits are seldom used in unconstrained environments, such as surveillance, as they require the cooperation of the individuals being observed during the acquisition process [21].

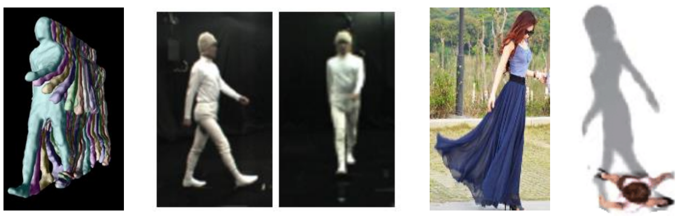
Gait is a behavioural trait, which, in contrast to most of the commonly used biometric traits, can be captured from a distance and even from poor resolution videos. The more significant advantage of a gait-based biometric recognition system is its non-cooperative operation, which allows the system to acquire gait features without the active participation of the individuals being observed, using acquisition systems such as a 2D camera. Since features describing gait are obtained from the rhythmic motion pattern, they are not easily prone to spoofing attacks or forgery, and are difficult to conceal, when compared to other biometric traits, in a daily life setting [19]. Such advantages make gait-based biometric recognition systems suitable for operation in unconstrained environments.

The performance of gait-based biometric recognition systems primarily depends on the quality of the features obtained. Thus, identifying novel gait representations that can best describe the gait of individuals can significantly improve the recognition results– see Figure 1.3 (a). When operating in unconstrained environments, other factors can affect the performance of the system – see Section 2.4, causing problems, such as:

- **Change in the viewpoint of the camera:** A 2D camera captures only the part of the 3D world visible to it. Thus, for a given camera viewpoint, parts of the individual's body will always be hidden from the camera. Due to the lack of cooperation from the individuals being observed, recognition systems may capture the gait from a viewpoint that is different from what is recorded in the database - see Figure 1.3 (b). Thus, although the captured gait features may correspond to a registered individual, they may differ from the representation stored in the database, leading to poor performance of the recognition system;

- **Change in the appearance of the individual:** In unconstrained environments, such as an area under surveillance, the appearance of individuals being observed can alter significantly due to changes in their clothing (e.g., by wearing a coat, a hat, or a long skirt) or by carrying items, such as a handbag or backpack. These changes can occlude parts of the individuals' body– see Figure 1.3 (c). The features obtained from the occluded parts may not match with the features recorded in the database, thus leading to poor recognition results.

An additional problem occurs when the viewpoint of the camera is changed such that the major articulations of the individuals' body cannot be observed – see Figure 1.3 (d). Under such conditions, the shadow cast by an individual can be an alternative source of gait features [22]. However, when captured in an unconstrained environment, the shadow may contain distortions and deformation caused by the camera, by a change in the type of the light source and by a change in the position of the light source with respect to the individual. Thus, performing recognition without rectification of the shadow silhouettes may lead to poor recognition results.



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

Figure 1.3: Illustrations of (a) a gait representation (b) change in viewpoint, (c) occlusion caused by change in appearance, (d) shadow cast by an individual [23].

1.1.2 Gait-based pathology classification

The way individuals walk is affected when the body systems that control gait do not function in the usual way the resulting gait is called pathological gait. It may happen due to illnesses, genetic factors, injuries or other problems, especially those affecting the legs or feet. The difficulty in walking may also be caused by acute problems, such as a bruise, cut, or fracture, but these temporary situations are not usually considered as gait pathologies, as the individuals in such cases can recover, often without medical intervention. Pathological gait can generally be categorised into one of five types, based on the symptoms or on the appearance of the individuals' gait [24], -see Figure 1.4:

- **Spastic gait:** It occurs when individuals drag their feet while walking. It can also make the individuals appear very stiff;
- **Scissor gait:** Individuals with scissor gaits will have their legs bend inward and they may even hit each other while walking. The crisscross motion resembles scissors opening and closing;
- **Steppage gait:** It occurs when individuals' toes point towards the ground while walking. Often, the toes will scrape against the ground as the individual steps forward;
- **Waddling gait:** Individuals with a waddling gait move from side to side when walking;
- **Propulsive gait:** It occurs when individuals walk with their head and neck pushed forward. It can appear as though the individual is rigidly holding a slouched position.

Such gait pathologies can be caused by several factors, such as injuries to the legs or feet, arthritis, infections in the soft tissue of the legs, broken bones in feet and legs, birth defects, shin splints, infections in the inner ear, cerebral palsy, stroke, tendonitis, conversion disorder or other psychological disorders. If the underlying conditions for the pathological gait are identified and treated, the gait of the individuals might be corrected. Even when medical treatment cannot correct the pathological gait completely, analysing the gait to identify the disorder can at least contribute to reduce the severity of the symptoms.

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

Figure 1.4: Illustration of (a) spastic gait, (b) scissors gait, (c) steppage gait, (d) waddling gait, (e) propulsive gait [25].

Pathological gait was traditionally diagnosed by medical professionals by asking the individuals questions about their medical history and symptoms, as well as by observing the way they walk. With the development of technology, gait analysis can now be performed using systems such as the motion capture system [18] or pressure mats [26]. Although such systems allow the collection of objective measurements for the analysis of the individuals' gait, they still require trained clinical personnel for their operation. These systems are also expensive to use, requiring specialised equipment and dedicated laboratories, therefore being inaccessible to most individuals in a daily life setting. The use of inertial sensors [27] can tackle some of the identified shortcomings, such as the need for dedicated laboratories and cost. However, these sensors need to be mounted on specific and distinct locations on the individuals' body, which can still only be done by trained clinical personnel. If one wishes to monitor individuals' gait in daily life settings in a non-invasive manner, systems that capture and analyse the gait using 2D cameras may be used [28]. However, to perform such analysis in the absence of clinical professionals, the systems should be able to capture features that best describes the gait pathology and analyses them automatically. The use of 2D vision-based systems involve additional problems, such as possible changes in the observation viewpoint, which coupled with the lack of depth information, can significantly affect the accuracy of the features obtained.

1.2 OBJECTIVES

This Thesis explores the use of gait observed by a 2D camera, to perform biometric recognition and pathology classification in unconstrained environments. Its objectives are:

- Reviewing the state-of-the-art systems available in the current literature;
- Proposing systems to address the limitations of the state-of-the-art;
- Capturing new datasets to evaluate the systems that operate on shadow silhouettes;
- Evaluating the performance of the proposed systems, in terms of accuracy, generalisation and complexity, as well as comparing them with the available state-of-the-art, while also ensuring the reproducibility of results.

This Thesis aims to improve the performance of the state-of-the-art systems by tackling the following problems:

- Identifying better gait representations, to improve the performance of recognition systems;
- Improving systems' robustness to changes in camera viewpoint and individuals' appearance;
- Exploring the use of shadow silhouettes in gait-based biometric recognition systems;
- Exploring the possibility of obtaining accurate biomechanical features that can be used to analyse gait impairments, even under a change in the viewpoint of the camera;
- Identifying features that can be used to classify gait across different gait related pathologies.

While examining the state-of-the-art systems that perform gait analysis, this Thesis also targets to propose novel encompassing taxonomies, whenever applicable.

1.3 CONTRIBUTIONS

This Thesis presents contributions that explore gait for biometric recognition and pathology classification in unconstrained environments. This work led to the following publications:

Journal

- T. Verlekar, H. Vroey, K. Claeys, H. Hallez, L. Soares, P. Correia. Robust Estimation and Validation of Biomedical Gait Indicators using 2D Video. Computer methods and programs in biomedicine, 175, 2019 [29];
- T. Verlekar, P. Correia, L. Soares. Automatic Classification of Gait Impairments Using a Markerless 2D Video-Based System. MDPI Sensors, 18(9), 2018 [1];
- T. Verlekar, L. Soares, P. Correia. Gait Recognition in the Wild using Shadow Silhouettes. Image and Vision Computing, 76, 2018 [30];
- T. Verlekar, P. Correia, L. Soares. View-Invariant Gait Recognition System Using a Gait Energy Image Decomposition Method. IET Biometrics, 6(4), 2017 [31].

International Conference

- T. Verlekar, P. Correia, L. Soares. Using transfer learning for classification of gait pathologies. IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, 2018 [32];
- T. Verlekar, P. Correia, L. Soares. Gait Recognition Using Normalised Shadows. European Signal Processing Conference, EUSIPCO, Kos Island, Greece, 2017 [33];
- T. Verlekar, P. Correia, L. Soares. Sparse Error Gait Image: A New Representation for Gait Recognition. International Workshop on Biometrics and Forensics, IWBF, Coventry, United Kingdom, 2017 [34];
- T. Verlekar, P. Correia, L. Soares. View-Invariant Gait Recognition exploiting Spatio-Temporal Information and a Dissimilarity Metric. International Conference. Of the Biometrics Special Interest Group, BIOSIG, Darmstadt, Germany, 2016 [35].

National Conference

- T. Verlekar, L. Soares, P. Correia. Shadow Type Identification for Gait Recognition using Shadows. 23th Portuguese Conference on Pattern Recognition, RECPAD, Lisbon, Portugal, 2017 [36].

In terms of biometric recognition, a novel taxonomy to organise the existing systems has been proposed [30]

Two novel gait representations have been proposed in the Thesis. The first, called the gait dissimilarity vector (GDV), consists of a feature vector of dissimilarity values, and it was published as referenced in [35]. It is obtained by computing the distances between an individual's gait energy image (GEI) and a set of GEIs called a prototype set, which is selected so that it is representative of the individuals registered in the database.

The second novel gait representation presented in this Thesis is called the sparse error gait image (SEGI), and it can be derived from the application of robust principal component analysis (RPCA) to a set of GEIs belonging to an individual. It highlights the most significant differences that exist among several gait sequences of an individual. SEGI explores the uniqueness of the resulting dissimilarities/differences for recognition [34].

Three novel systems to detect the viewpoint of the camera are proposed in this Thesis. By detecting the viewpoint, the recognition systems can limit the matching module to a subset of the database that corresponds to the identified viewpoint, thus improving their recognition accuracy. The first system, detects the viewpoint of the camera by obtaining features, representing the viewpoint, from the leg region of the observed individuals, followed by matching those features with the database. The gait of the individuals is represented using the GEI and the features representing the overall structure of the leg region are obtained using a perceptual hash (PHash) function [31].

Since, the above viewpoint detection system can only detect the viewpoints it is trained for, a second system has been developed, performing view detection without any training [35]. This system operates in three steps:

- It identifies the feet position of the individuals along time using a gait representation called the gait texture image (GTI);
- The feet positions are used to estimate the direction of the dominant walking trajectory;
- The viewpoint is determined by associating it to the identified walking trajectory direction.

Since neither the PHash nor the GTI based walking viewpoint detection systems can operate in the presence of shadows under the feet, a third system is presented in this Thesis [30], [33]. This system analyses the orientation of the individuals' body and the corresponding shadow to detect the feet position. The viewpoint can then be estimated by processing the evolution of feet positions over time, following the steps presented in the second system. Since, the shadow connects to the body at the feet position; the proposed system can also be used to segment the shadow silhouettes.

The problem of appearance change is addressed by presenting a gait-based biometric recognition system that can identify and discard the altered sections of the individual's GEI being observed [31]. It operates in four steps:

- The gait of the individual is represented using a GEI;
- The GEI is decomposed into sections;
- The altered sections are identified by comparing the GEI to the normal appearance of the individuals in the database using a representation called the average GEI image;

- Altered sections are discarded and the recognition is performed by aggregating the votes from the remaining sections.

Thus, the system performs well regardless of the type of appearance change encountered.

The possibility of using shadow silhouettes as an alternative source of gait features for biometric recognition is also explored in this Thesis. To use the shadow silhouettes, they must first be rectified, compensating the distortions and deformations caused by the camera perspective and parameters. Two novel systems are presented to rectify the shadow silhouettes into a common canonical view, using a transformation matrix. The first system obtains a transformation matrix by applying the transform invariant low-rank texture (TILT) to a GTI gait representation [33]. The second novel system obtains the transformation matrix using a 4-point correspondence system. It identifies the head and feet position of the individual in the first and final frames of the video sequence and estimates their location in the canonical viewpoint. The system obtains the transformation matrix using these positions [30]. The two systems perform recognition using GEIs obtained from the rectified shadow silhouettes. To evaluate the performance of the systems the Thesis also presents a new dataset, consisting of 21 individuals moving along two walking directions [30] captured at Instituto Superior Técnico (IST) in Portugal.

A limitation of the gait-based biometric recognition systems that rely on shadow silhouettes is that they operate only on shadow silhouettes that appear similar to the individuals' body. However, the shadow cast by an individual may be useful for recognition or appear as a blob of undefined shape depending on the lighting conditions - see Section 2.4.3. Thus, a novel system is presented in this Thesis that analyses the intensity ratio between the static background and the dynamic foreground comprising of the individuals and their shadow, to check if the shadow is usable [36]. To evaluate its performance a dataset is also captured containing six individuals walking on two different days casting two different type of shadows.

The second area in which this Thesis contributes is in the usage of gait analysis for detection of accurate biomechanical features and pathology classification.

A novel system is presented to accurately estimate various temporal gait features, such as stance time, swing time and gait cycle time, by detecting for each gait cycle the moments of initial contact of a foot with the ground and when the foot loses contact with the ground, known as the toe off event, using a single 2D camera [29]. The proposed system maintains the high level of accuracy, even when the viewpoint of the camera changes, making it suitable for clinical evaluations. To validate its use in clinical evaluations the system is compared with the output of the motion capture system [18], considered as the gold standard for the medical analysis of gait. To perform the comparisons a new dataset has been captured in cooperation with the Rehabilitation Sciences group of the Katholieke Universiteit Leuven (KUL), Bruges campus, in Belgium. The acquired dataset contains data from 10 individuals, recorded using both a 2D camera and the motion capture system installed in KUL's gait analysis lab [18].

The Thesis presents two novel systems to perform classification of gait across different gait related pathologies. The first system explores two different types of features to perform pathology classification:

- Feet related features, including step length, step length symmetry, fraction of foot flat during stance phase, normalised step count and speed. These features help classify gait pathologies that affect the individuals' feet;

- Body related features – since some gait related pathologies also affect other body areas and the body posture, body related features include the amount of movement while walking, shift in the centre of gravity and torso orientation.

Together, feet and body related features allow performing classification of the observed walking sequences across different gait related pathologies [1].

The final system presented in this Thesis, explores the popular GEI biometric gait representation together with the usage of a deep convolutional neural network (CNN), to perform pathology classification [32]. The use of GEI makes the system robust to silhouette segmentation errors, which can severely affect the classification results. The performance of the system is further improved by using a fine-tuned VGG-19 convolutional neural network for feature extraction. The fine-tuning process, using a transfer learning strategy, allows the system to extract features that best represent the gait related pathologies, without the need of a very large training set.

1.4 THESIS STRUCTURE

This Thesis is organized into four parts. Part I discusses gait analysis basic concepts and the state-of-the-art, containing three chapters. The first chapter (i.e., this chapter) introduces the Thesis presenting the context, motivation, objectives and contributions, along with a list of publications achieved during its preparation. Chapter 2 begins with a discussion about basic concepts of gait, followed by the presentation of various acquisition systems and a summary of challenges faced by 2D vision-based systems when used for gait analysis in unconstrained environments. The last chapter of Part I presents the current state-of-the-art systems, highlighting their advantages and disadvantages when operating in unconstrained environments.

Part II of the Thesis comprises Chapters 4 to 7 and presents the biometric recognition contributions, along with their performance evaluation. The proposed GDV and SEGI gait representations are described in the fourth chapter. Chapter 5 presents three novel systems proposed to tackle the problem of change in viewpoint of the camera, by using minimal information from the leg region. Chapter 6 presents a system that tackles appearance change by decomposing GEIs into sections. The final chapter of Part II explores the use of shadow silhouettes for gait-based biometric recognition. It presents two novel systems for the rectification of shadow silhouettes and a system to identify the type of shadow cast by the individuals.

Part III of the Thesis presents novel contributions for medical diagnosis, exploring the use of gait in clinical analysis and classification of gait related pathologies, along with their performance evaluation. A novel system is proposed in Chapter 8 that obtains temporal gait features with a high level of accuracy, even when the viewpoint of the camera changes. The Thesis then presents two novel systems to perform classification of gait across different gait related pathologies, in Chapter 9. The first system explores the use of biomechanical features for classification, while the second system presents the use of biometric gait representations along with a deep CNN.

Finally, Part IV concludes the Thesis, with Chapter 10 containing a summary of achievements along with possible future research directions.

2 GAIT ANALYSIS: BASIC CONCEPTS AND ACQUISITION SYSTEMS

2.1 INTRODUCTION

This chapter discusses some basic concepts related to gait representation, presents systems that can be used to acquire gait information, identifies a set of challenges faced by 2D vision-based gait analysis systems when operating in unconstrained environments, and concludes by presenting the general architecture of a gait analysis system and the evaluation metrics usually adopted in the literature.

2.2 GAIT STRUCTURE AND REPRESENTATION

The gait of an individual refers to the manner of walking, rather than the actual walking process [37]. Its analysis primarily involves observing the structured movement of the lower limbs, which repeat a series of phases constituting a gait cycle. Considered the functional unit of gait, a gait cycle can be divided into the stance and the swing phases [38].

The stance phase, as illustrated in Figure 2.1, is the part of a gait cycle that occurs when the foot being observed (highlighted right foot in Figure 2.1) is in contact with the ground. It begins with an event called the initial contact (IC) or the heel strike and ends with an event called the toe off (TO) of same foot. The stance phase accounts for around 60% of the duration of the entire gait cycle. The remaining 40% of the duration is composed of the swing phase, which is the part of the gait cycle that occurs when the observed foot is not in contact with the ground. It begins when the observed foot first leaves the ground and ends with the heel strike of the same foot; it is the non-weight bearing phase of the gait cycle.

Alternatively, the gait cycle can also be divided into 2 double support and two single support phases [38], as illustrated in Figure 2.1. The double support phase covers the duration of the gait cycle when both feet of the individual are in contact with the ground. It begins when one limb ends the stance phase while the other limb begins it. About only 10 percent of the gait cycle is spent in the double support phases. The rest of the gait cycle, where only a single foot is in contact with the ground, is defined as the single support phase. The percentage of time spent in each phase changes with a change in the speed of an individual's movement. Slower speeds lead to more percentage of time in the double support phase, while faster speeds lead to less percentage of time in the double support phase. The transitions from walking to running, is indicated by the addition of a third phase, called the non-support phase, where none of the feet are in contact with the ground.

The gait cycle contains a series of events that correspond to very specific instants in each phase. The stance phase of the gait cycle includes five events and the swing phase includes three events, as illustrated in Figure 2.1. They can be defined as [39]:

- **Initial contact/Heel strike:** It occurs when the heel of the leg being observed first meets the ground, as illustrated by the highlighted (right) leg in Figure 2.1. It is indicated by the ankle of the leg being observed in neutral position, slightly flexed knee and approximately 30 degrees flexion of the hip with respect to the vertical. With the heel strike, the body weight of an individual begins to shift onto the leg being observed;

- **Loading response/Foot flat:** It occurs when the observed foot is in complete contact with ground. It is indicated by an approximately 5 - 10 degrees extension of the ankle of the leg being observed, and 15 degrees flexion of the knee. The hip is extended, allowing the trunk and the body to catch up with the leg, while the weight of the body continues to shift onto the leg being observed;
- **Mid-stance:** It is the instant during the gait cycle in which the body passes over the leg being observed. The leg approaches a vertical position, offering a single-leg support, with the other leg freely swinging forward. The ankle flexes, the knee extends while the hip of the leg being observed continues extending. The trunk enters a neutral position where the arms are parallel to the body;
- **Terminal stance/Heel off:** It occurs when the heel of the leg being observed begins to lift off the ground. The ankle and the knee extend, the hip hyperextends, and the trunk rotates to the side of the leg being observed;
- **Pre-swing:** It occurs at around 60 % of the duration of the gait cycle and marks the end of the stance phase. It corresponds to the start of the gait cycle's second period of double support in which the body weight is transferred from leg being observed to the other;
- **Toe off:** It is the instant when the toes of the leg being observed just leave the ground while the other leg is in full contact with the ground. It indicates the end of the stance phase and beginning of the swing phase. During the toe off (TO), the toes of the leg being observed go into hyperextension, the ankle extends, while the knee and the hip start flexing;
- **Mid-swing:** It involves shortening of the leg to clear the ground. The ankle of the leg being observed is in neutral position, the knee in maximum flexion, while the hip continues flexing;
- **Terminal swing:** The knee joint of the leg being observed extends in the preparation for the stance phase.

These events can be used to estimate various kinematic features useful to describe an individual's gait. Accurate and reliable knowledge of these kinematic features and their evolution over time can help in the diagnosis of diseases, recognition of the walking person's identity and the analysis of the different forces exerted on muscles, which can be useful for instance in the field of sports. These kinematic features can be divided into three different groups: spatial, temporal and spatiotemporal [2].

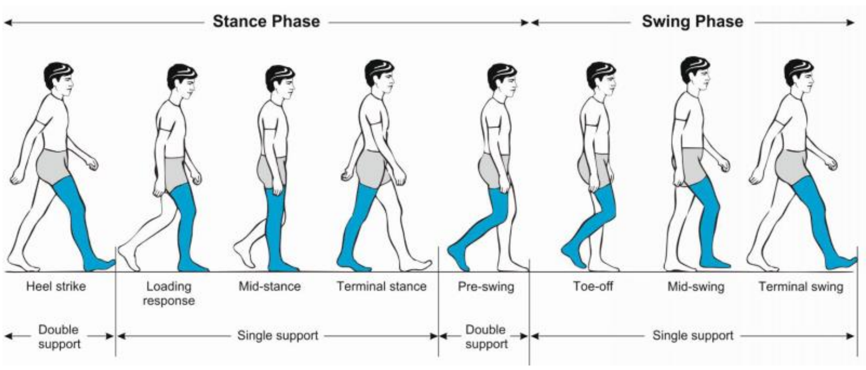


Figure 2.1: Key events and phases that occur during a gait cycle [39].

The spatial group includes distance and angular features, as illustrated in Figure 2.2, such as:

- **Stride length:** It is the distance travelled in one stride, i.e., between two consecutive heel strikes of the leg being observed, which constitutes a gait cycle;
- **Step length:** It is the distance covered between heel strike of the leg being observed and heel strike of the other leg;
- **Step width:** It is the distance by which two feet are spread apart during walking. It is measured as the lateral distance from the heel centre of the leg being observed to the line of progression formed by two consecutive foot contacts of the other foot;
- **Degree of toe out:** It represents the angle formed by each foot's line of progression and a line intersecting the centre of the heel and the index toe;
- **Joint angles:** They represent the angles formed by different joints of the body, especially the ankle, the knee and the hip of the leg being observed.

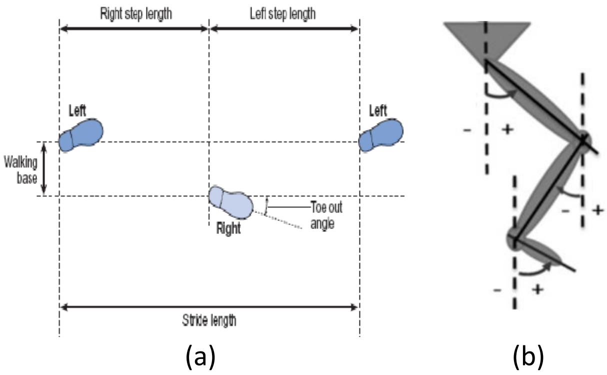


Figure 2.2: Illustration of spatial features, including (a) distance and (b) angular features.

The temporal group includes time-related features, such as:

- **Stride time:** It is the time required to complete one gait cycle, also called period of a gait cycle;
- **Step time:** It is the time covered between the heel strike of the leg being observed and the heel strike of the other foot;
- **Stance time:** It is the amount of time spent in the stance phase;
- **Swing time:** It is the amount of time spent in the swing phase;
- **Single limb time:** It is the amount of time spent with only a single foot in contact with the ground during a gait cycle;
- **Double limb time:** It is the amount of time spent during a gait cycle, with both feet in contact with the ground.

The spatiotemporal group includes features-related to both the spatial and temporal dimensions, such as:

- **Cadence:** It is the total number of steps completed within a given period;
- **Speed:** It is the distance covered per unit of time.

Apart from kinematic features, several other features can contribute to describe an individual's gait. They include:

- Ability of an individual to walk without external support;
- Existence of tremors while walking;

- Record of falls;
- Ground reaction forces exerted by the foot while walking;
- Electrical activity produced by muscles while walking;
- Posture of the body, which includes bending of the spine and symmetry between the two sides while walking.

Traditionally, walking individuals were observed by specialists to perform gait analysis, possibly including the estimation of a set of representative gait features. The subjective assessment of the individuals performed by the specialists included tasks such as observing the time taken by the individuals to get up from a sitting position, walking certain distances, turning around and sitting back down again. It also included observing and recording various gait features such as speed, step length and balance. Apart from visual cues, the assessment also included conducting interviews to obtain information about the individuals' gait.

2.3 GAIT ACQUISITION SYSTEMS

Since a gait cycle involves several important motions that can be missed by a naked eye observation, various automatic acquisition devices are now routinely used to capture gait information and allow a subsequent more objective gait analysis.
The various systems available for gait acquisition and analysis can be broadly classified into systems with wearable or non-wearable sensors [2], following the proposed taxonomy illustrated in Figure 2.3 and discussed in the following subsections.

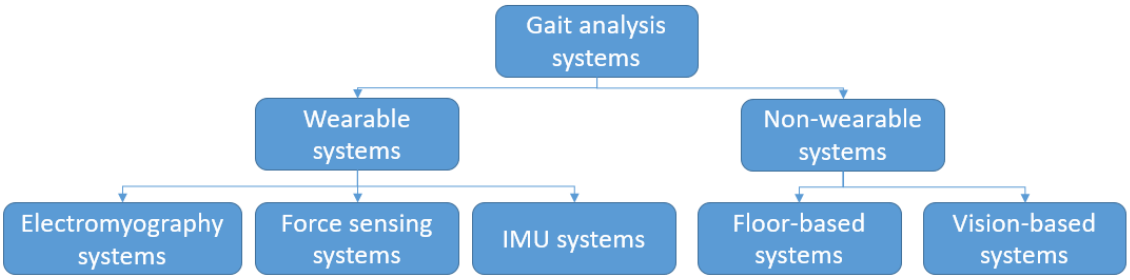


Figure 2.3: Proposed taxonomy for the state-of-the-art gait analysis systems.

2.3.1 Systems with wearable sensors

Wearable sensors can be attached to various parts of the body to acquire features that can describe an individual's gait. Systems that use such sensors include:

- **Force sensing systems**: These systems use sensors, such as force sensitive resistors [40], which measure the ground reaction forces under the foot as the potential difference proportional to the measured pressure. Such sensors are typically attached to the insole of the shoes, as illustrated in Figure 2.4 (a). As the weight placed on the sensors increase, their resistance decreases, thus providing the change in potential;
- **Inertial measurement unit (IMU) systems**: These systems use sensors, such as accelerometers [27] and gyroscopes [41] , called the IMUs, which measure acceleration, velocity and orientation using the laws of motion - see Figure 2.4 (b). Accelerometers estimate the acceleration by measuring the net force acting on them. They can also estimate the angular velocities and the flexion angles from the resulting acceleration. The gyroscope tracks rotation, twist and change in orientation using rotational inertia. The resistance to change in rotation speed and turn direction produces a potential difference from which such values are sensed;

- **Electromyography systems**: Electromyograph [42] is another type of wearable sensor that detects the electrical signals resulting from contracting muscles - see Figure 2.4 (c). It can be used to measure the relative muscle tension caused by increased walking speed or the initiation of the various events of the gait cycle.

(a) (b) (c)

Figure 2.4: Examples of wearable sensors: (a) force sensitive resistors [43]*, (b) inertial measurement units* [44]*, (c) electromyograph* [45]*.*

There are several advantages to using wearable sensor-based systems to analyse an individual's gait. They are relatively cheap, portable and can transfer captured signals wirelessly or through memory transfer, allowing them to acquire gait over long periods. Such systems can also operate in uncontrolled environments, thus extending their use beyond special laboratories. These advantages allow the wearable systems to be used for daily and long-term monitoring. However, setting up an individual with such systems requires clinical professionals, as the position of the sensors on the body of the individuals must be precise. Clinical professionals are also required to interpret correctly the signals acquired from these sensors, which are also susceptible to noise and interference from external factors. Additionally, most of them can acquire only a limited number of gait features, thus limiting a complete analysis of individuals' gait.

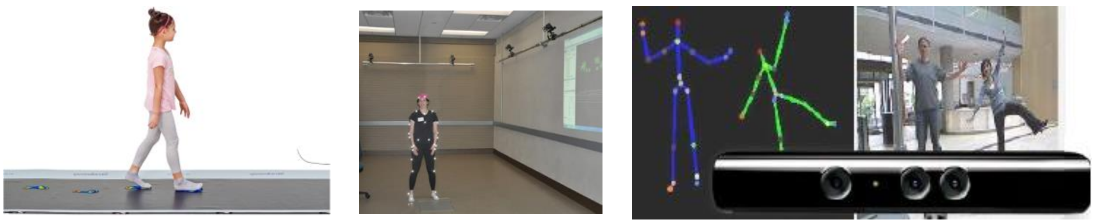
2.3.2 Systems with non-wearable sensors

Non-wearable sensor-based systems acquire gait features without physically attaching the sensors to the individual's body. Depending on the type of sensors used, they can be further divided into:

- **Floor-based systems**: They involve the use of sensors such as pressure mats [26] that quantify the pressure patterns under the feet. They also include the use of force platforms [46], which quantify the horizontal and shear components of the forces applied, along with the centre of pressure. These systems can be setup on the floor to capture information about the gait of individuals walking over it, as illustrated in Figure 2.5 (a);
- **Vision-based systems:** These systems acquire images using one or multiple optical sensors. The resulting images can then be processed to obtain the desired gait features. Vision-based systems can be further grouped into marker-based and marker-less systems:
    - **Marker-based systems:** The most widely used systems in medical environments, being considered the gold standard in clinical analysis, are marker-based systems, such as the motion capture system [18], which relies on the application of multiple markers to various parts of the body, together with a setup containing multiple calibrated optical sensors to capture the individuals' gait - see Figure 2.5(b). The markers are attached to key body positions that can best characterise individuals' gait. The calibrated optical sensors then capture the light reflected/emitted by the markers to estimate its coordinates in 3D space.

- **Marker-less systems:** Marker-less systems can use depth-sensing cameras [47], or multiple 2D cameras [48], to capture individuals' gait. Recently, a significant amount of work has also been done considering marker-less systems using a single 2D camera. The systems that rely on depth-sensing cameras project a known pattern onto the scene using for instance infrared laser light and infer the depth from the deformation of that pattern. The deformations, which indicate the distance the infrared light travels, are captured by an IR sensor to construct a depth map. They then estimate the body positions in 3D space using the depth map - see Figure 2.5 (c). The usage of multiple 2D cameras allows a system to capture images of individuals from multiple viewpoints. Every image can be used to estimate the 2D coordinates of the key joint positions from a different viewpoint. The systems can then combine those coordinates to obtain the joint positions in the 3D space, using the principles of a stereo camera system.

Floor-based systems are non-intrusive in terms of attaching sensors to the body. However, the gait information acquired by these systems, such as the ground reaction forces, is often relatively basic and can only be used to perform a general analysis of the individuals' gait. Another limitation of this type of systems is that they can only operate in controlled environments, such as laboratories, where the walking path is clearly defined.



(a) (b) (c)

Figure 2.5: Non-wearable sensor-based systems: (a) floor-based system [49], (b) marker-based vision system [50], (c) marker-less vision system [51].

The marker-based vision systems can acquire accurate and precise gait features from the 3D model they build for the individuals. However, the use of marker-based systems is also limited to special laboratories as the cameras are sensitive to illumination conditions and require calibration before use. This type of system is partially invasive, as the markers must be physically attached to the body of the individuals. In addition, the placement of the markers must be precise which necessitates the help of clinical professionals.

The marker-less systems are popular for their non-invasive acquisition of gait. However, the accuracy of the estimated body positions is typically much lower than for the gold standard marker-based motion acquisition systems. In addition, depth-sensing cameras typically operate with a limited range between 80 cm and 4 m and the usage of multiple cameras requires calibration before use.

Major articulations during a gait cycle occur in the sagittal plane [52]. Thus, 2D marker-less vision-based systems usually rely on a side view observation of an individual to perform gait analysis. To obtain gait features, the captured images are typically converted into binary silhouettes using background subtraction [53] - see Figure 2.6 (a). The silhouettes that highlight the walking individuals can then be used to estimate various gait features such as step length, leg angles, gait cycle time [54], cadence, speed, and stride length [55], or the fraction of the stance and swing phases during a gait cycle [56] - see Figure 2.6 (b). A drawback of such systems is that they do not have access to depth information, which limits their accuracy when compared to other vision-based systems. However, they provide several advantages for gait acquisition. The number of features estimated by such systems is significantly larger than most wearable or floor-based systems. Being non-invasive, they do not require the use of markers on the individual's body, do not require calibrations, and are robust to lighting conditions, making them suitable for usage also in unconstrained environments. They can be deployed in a wide range of applications, from surveillance of large spaces to gait analysis in clinics. They are also inexpensive, accessible and can be used by normal individuals in a daily life setting. Thus, due to their advantages over other acquisition systems, this Thesis considers a single 2D camera to perform gait analysis in unconstrained environments.

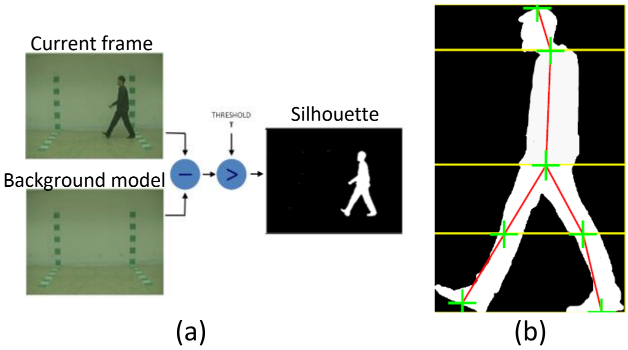


Figure 2.6: Steps in 2D vision-based gait analysis (a) background subtraction, (b) feature estimation.

2.4 CHALLENGES TO 2D VISION-BASED GAIT ANALYSIS IN UNCONSTRAINED ENVIRONMENTS

2D vision-based systems can be used to analyse the gait of individuals to recognise their identity or detect whether their gait is affected by some pathology. Performing such analysis in a constrained environment provides several advantages such as active participation of the individuals, lighting control, calibrated systems, predefined walking path and speed, absence of occlusions, and appearance control. Such advantages are lost when gait is captured in an unconstrained environment. Under such conditions, the performance of 2D vision-based systems depends on many factors. To allow a better understanding of the relevant factors, a novel taxonomy, is presented in this Thesis – see Figure 2.7, grouping these factors into four main dimensions:

- Individual-related factors;
- Camera-related factors;
- Light source-related factors; and
- Environment-related factors.

Some of the considered factors affect the individuals' physical gait, while others affect the observation of gait and, therefore, the possibility to capture good features. Each of these dimensions, as well as some combinations of factors from different dimensions, is discussed in the following subsections.

2.4.1 Individual-related factors

To perform the analysis of individuals' gait, most 2D vision-based systems rely on features describing the individuals' appearance and the corresponding motion pattern while walking. As such, factors that affect the observable features or the physical gait of the individuals should be taken into account. One such set of factors is related to the individuals' appearance. Most systems that operate in a constrained environment assume that the individuals' appearance remains the same along time [57]. However, this is not true in an unconstrained environment where the individuals' appearance can be altered by clothing changes, e.g. when wearing a coat, a hat, or a long skirt, as well as by carrying items, such as a bag or backpack. This can result in (partial) occlusion of the gait features, as can be seen by comparing Figure 2.8 (a) and Figure 2.8 (b).

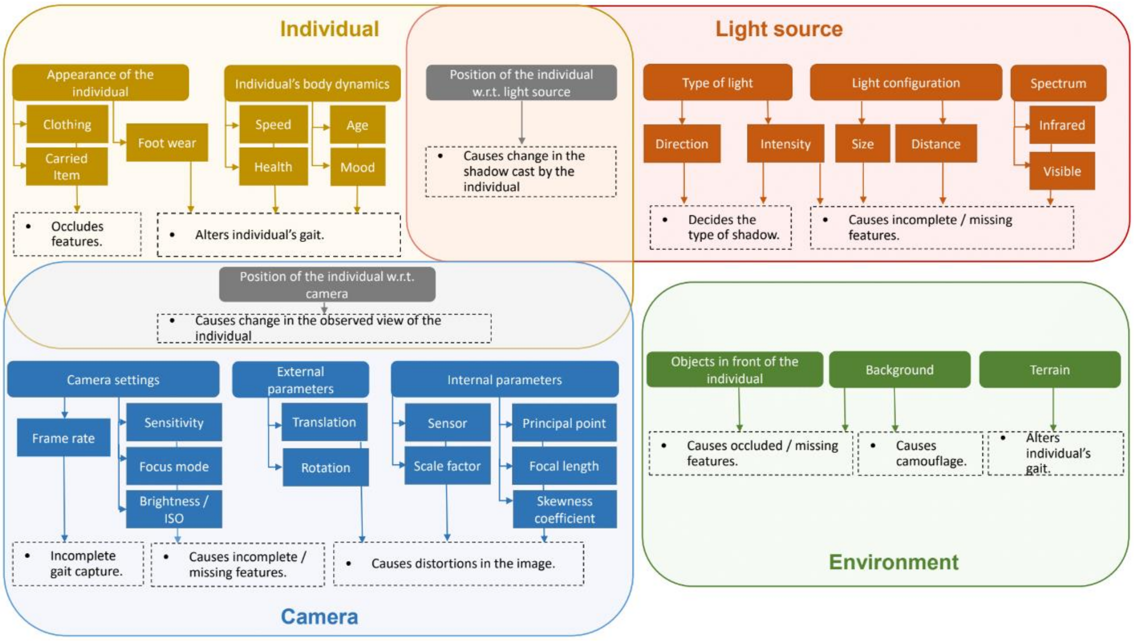


Figure 2.7: Factors affecting gait analysis using 2D vision-based systems.

A second set of factors affect the individuals' body dynamics, thus changing their gait. It includes factors such as speed, health, age and mood of the individuals. An increase in the walking speed of the individuals can cause changes to their gait, such as increase in the arms swing and the stride length, as well as torso orientation, as illustrated in Figure 2.8 (c). Similarly, injuries or other health issues may cause shorter or irregular strides, bending of the spine, or restricted limb movement, altering the individuals' gait, as illustrated in Figure 2.8 (d). Age is another factor that affects individuals' gait as the comparisons made between features acquired over longer periods of time are prone to be affected by changes in the individuals' height, weight, and muscular and bone development.

Other factors that can cause changes to individuals' gait include footwear, and individuals' mood. The impact of using different footwear has been addressed in the literature, for instance [58], and the influence of mood on gait has been discussed in the medical literature, such as [59].

2.4.2 Camera-related factors

2D vision-based systems estimate features from images captured by a 2D camera. Therefore, any image distortions and deformations caused by the camera system, typically described by its external and internal parameters [60], can affect the computation of the gait features. Camera external parameters include rotation and translation matrices, denoting the transformations from the 3D world coordinate to the 3D camera coordinates. Camera internal parameters include focal length, image sensor format, principal point, scale factor and skewness coefficient, which defines the transformation of 3D camera coordinates into a 2D image (following a pinhole camera model). Thus, the combination of internal and external parameters describes the transformation of the 3D world coordinates into a 2D image. The internal and external parameters also control the field of view, scale, skewness and resolution of the resulting image, with any distortions or deformations in the image affecting the captured gait features [33], as illustrated in Figure 2.9.

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

Figure 2.8: Examples of silhouettes for: (a) normal gait, (b) gait partly occluded by a bag, (c) gait altered by speed and (d) gait affected by health.

Estimation of good gait features depends on the successful capture of key events of a gait cycle, as illustrated in Figure 2.1. Since a normal gait cycle lasts for only a few seconds, the camera must capture images at a sufficiently high frame rate to capture these key events (accurately). Therefore, the acquisition frame rate of the camera is a determinant factor for the quality of the obtained gait features [61]. Other camera-related factors are the sensitivity, focus mode and brightness/white balance control settings, which can help to better distinguish the individuals from the background, thus affecting the quality of silhouettes from which features are often extracted. Poor camera settings may lead to incomplete or missing features. As discussed in [62], changes in silhouette quality (i.e., incomplete features) can affect the performance of the 2D vision-based system.

Figure 2.9: Examples of shadow silhouette deformations caused due to field of view, skewness and scale changes.

2.4.3 Light source-related factors

The third taxonomy dimension is related to the light source illuminating the scene. Light intensity is the main factor determining whether it is possible for the camera to "see" and therefore acquire gait features, or if the scene is too poorly illuminated thus preventing the acquisition of usable features. Another important factor that characterises the light source is its spectrum range. For instance, the works reported in [63] and [64] discuss the merits of using infrared light to perform gait analysis.

The scene illumination also determines whether there will be a shadow cast by the individuals, which can be used for analysing their gait [33]. The quality of the shadow silhouette depends primarily on the direction of the light rays emitted by the light source. When light rays travel in the same, well-defined direction, towards the individuals, i.e., under a collimate source of light, a sharp shadow is cast, displaying features similar to those that can be computed from the individuals' body silhouette – see Figure 2.10 (a). If the light rays have many different directions, the resulting shadow will be diffused, appearing as a blob around the individuals with no distinguishing gait features – see Figure 2.10 (b).

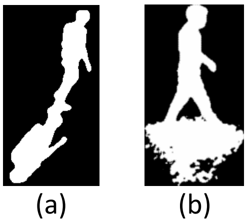


Figure 2.10: Examples of (a) sharp and (b) diffused shadows cast by the walking individual.

Other factors affecting the cast shadows are the distance and size of the light source. If the individuals are too close to the light source, an incomplete shadow will be produced, only including the body parts illuminated by it. On the other hand, if the individuals are too far away, a sufficient intensity of light may not reach them to cast a usable shadow. Finally, the illumination source size should be sufficiently large to illuminate uniformly the entire scene. However, if the light source size is large but not sufficiently far away then light rays from multiple directions might cause multiple overlapping shadow contributions, resulting in a diffused shadow from which gait features are not identifiable.

2.4.4 Environment-related factors

The environment, as observed by the camera, contains three main factors that can affect gait analysis:

- **Objects in front of the individuals**: Objects in front of individuals can cause occlusions and make it impossible for the camera to capture all the necessary body parts, which can lead to missing features;
- **Background properties:** Some properties of the background, such as colour and texture, if similar to the individuals' clothing, can cause camouflage. In this case, distinguishing the individuals from the background becomes difficult, which can also lead to incomplete or missing features;
- **Terrain on which individuals walk**: The terrain on which the individuals are walking can cause changes to their gait. Those changes can be attributed to terrain properties, such as elevation, friction and irregularities. For example, individuals have to put extra efforts to walk up a slope, when compared to a flat surface. These efforts, as well as terrain properties, alter their arm swing, stride length and the orientation of some body parts. Thus, the terrain can significantly affect the observed features, as discussed in [65].

2.4.5 Combination of factors

The above taxonomy dimensions help to understand the factors affecting 2D vision-based systems. However, in unconstrained environments, some factors appear in combination across different dimensions, with further effects on the observed gait features. For instance, a camera captures only the part of the 3D world visible to it. Therefore, changes in the individuals' position and walking direction relative to the camera result in a change in the viewpoint as observed by the camera. Under such conditions, the features captured by the camera can change even when the physical gait of the individuals remains unchanged and un-occluded. For example, when individuals walk towards the camera, their front view is observed, as illustrated in Figure 2.11 (a), while if they walk perpendicularly to the camera axis, the observed side view – see Figure 2.11 (b) – is significantly different. Even when they walk along a straight line across the (fixed) camera, the captured view at the start of the gait sequence – Figure 2.11 (c) – will be significantly different from the one at the end – see Figure 2.11 (d). Consequently, features computed for the same walking direction with a fixed camera may not match each other.

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

Figure 2.11: Different viewpoints of the same individual: (a) frontal view, (b) side view, (c) view at the start of the sequence, (d) view at the end.

Another relevant combination across taxonomy dimensions, especially when considering the use of shadows for gait analysis, is that of individuals and light source factors. This is especially relevant when there are considerable body self-occlusions while the cast shadow is clearly visible, for instance, when a camera observes individuals from the top view. In this case, the cast shadows depend on the combination of their walking direction and the illumination direction, as illustrated in Figure 2.12 (a, b).

(a) (b) (c)

Figure 2.12: Change in the shadow cast by the same individual: (a, b) when the individual's position w.r.t. the light source changes, and at the (b) start and (c) end of a linear walking sequence.

In addition, depending on the relative positions of the individuals, light source and camera, the individuals' body may occlude the cast shadow. However, one interesting observation is that, under a distant collimated light source, when the individuals walk along a straight line perpendicular to the camera axis, the shadows cast by them along the trajectory will always appear similar to each other, since their position with respect to the light source does not change significantly. This is different from what happens for body silhouettes, as illustrated in Figure 2.12, where the shadow silhouette (b) appears to be a skewed version of the shadow silhouette (c). In addition, since the cast shadow corresponds to an area darker than its surroundings, it can be more easily distinguished from the background using background subtraction [66].

2.5 SYSTEM ARCHITECTURE AND EVALUATION METRICS

The general architecture of a 2D vision-based system for gait analysis, as illustrated in Figure 2.13, can be divided into 3 main modules:

- **Pre-processing**: The pre-processing module converts an input obtained during acquisition into a usable sequence that can be processed by the system.
- **Feature extraction:** The feature extraction module converts the results of the initial pre-processing into a gait representation, which is capable of capturing individuals' distinctive gait characteristics. The gait representation can itself be used as a feature for matching or additional features can be extracted from it.
- **Matching:** The final module matches the features obtained from an input to the features recorded in the database during enrolment using a classifier. The matching module in the case of a gait-based biometric recognition system reveals the identity of the individuals, while in the case of gait-based pathology classification the system reveals the gait pathology the individuals are suffering from.

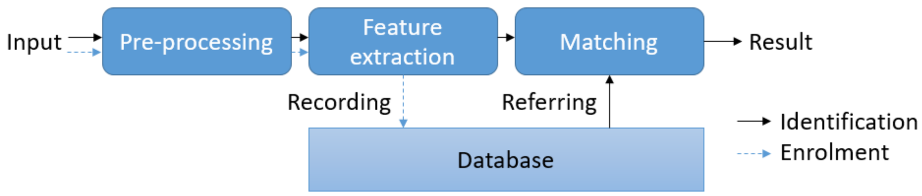


Figure 2.13: Architecture of the 2D vision-based systems for biometric recognition and pathology classification.

The performance of a gait analysis system can be improved by improving the 3 modules. However, most state-of-the-art systems, discussed in Chapter 3, do not focus on pre-processing, as it is better addressed in other domains of computer vision, such as foreground detection [66] and video denoising [67]. The state-of-the-art systems improve their performance by enhancing the quality of the gait representations used and by using better performing classifiers.

The performance of these systems can be visually represented using a confusion matrix. It is a specific table layout where each row presents the predictions while each column represents the ground truth. With respect to each entry in the confusion matrix, the table can be divided into 4 quadrants as follows:

- **True positive (TP):** It is a correct positive prediction;
- **True negative (TN):** It is a correct negative prediction;
- **False positive (FP):** It is a false positive prediction, also called as type 1 error;
- **False negative (FN):** It is a false negative prediction.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="4">Ground truth</td></tr>
<tr><td>1</td><td>2</td><td>…</td><td>N</td></tr>
<tr><td rowspan="4">Prediction</td><td>1</td><td>TP</td><td colspan="3">FP</td></tr>
<tr><td>2</td><td rowspan="3">FN</td><td colspan="3" rowspan="3">TN</td></tr>
<tr><td>…</td></tr>
<tr><td>N</td></tr>
</table>

Figure 2.14: Illustration of a confusion matrix

The confusion matrix can be used to evaluate the performance of the system by computing the (recognition/classification) accuracy. It is defined as the percentage of correct classifications among the total number of classifications performed. It can be computed according to (1.1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1.1}$$

The confusion matrix can be used to obtain other evaluation metrics, such as recall, precision and specificity of the system. Recall is also referred to as the true positive rate or the sensitivity of the system. It represents the percentage of correct classifications among the total number of individuals present in a group. It can also be interpreted as the probability that a (randomly selected) relevant individual is retrieved in a search. It can be computed according to (1.2).

$$Recall = \frac{TP}{TP + FN} \tag{1.2}$$

Precision, which is also referred to as positive predictive value, represents the percentage of correct classifications among the total number of individuals classified. It also represents the probability that a (randomly selected) retrieved individual is relevant. It can be computed according to (1.3).

$$Precision = \frac{TP}{TP + FP} \tag{1.3}$$

Specificity measures the proportion of actual negatives that are correctly identified as such. Also called the true negative rate, it can be measured according to (1.4).

$$Specificity \ = \frac{TN}{TN + FP} \tag{1.4}$$

Another possible way of visually representing the performance of the system is using the receiver-operating characteristic (ROC) or the cumulative match characteristic (CMC) curves – see Figure 2.15. ROC is a probability curve plotting recall against 1-specificity. The area under the curve represents the degree or measure of separability, i.e. the capability of the system in distinguishing between individuals/pathologies. The higher the area under the curve, the better the accuracy of the system. The CMC is a precision curve that provides the accuracy of the system for each rank, where the rank is defined by the number of queries made to the database to obtain the correct result. CMC is plotted with accuracy (classification/recognition rate) against the rank.

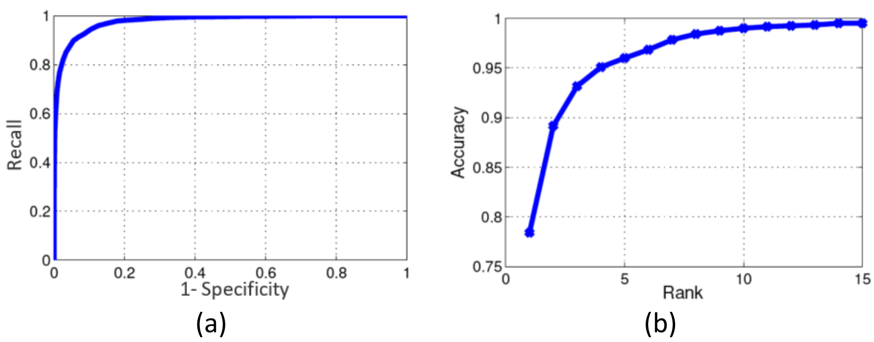


Figure 2.15: Illustration of a (a) ROC and (b) CMC curve.

3 GAIT ANALYSIS: STATE-OF-THE-ART

3.1 INTRODUCTION

Gait analysis can be a challenging task, especially in unconstrained environments, where certain factors or the combination thereof can cause problems, which are perceived as change in the viewpoint of the camera or change in the appearance of the individuals being observed – see Section 2.4. This Thesis presents solutions to some of the problems faced when using gait analysis in 2D vision-based biometric recognition and pathology classification systems, operating in unconstrained environments. However, to emphasise the significance of the proposed systems they must be compared to the state-of-the-art 2D vision-based systems available in the current literature. This chapter provides a review of the current state-of-the-art, addressing the problems of:

- **Gait-based biometric recognition**: The review of gait-based biometric recognition systems includes a discussion on various available gait representations, existing systems that are robust to viewpoint and appearance changes and systems that perform recognition using shadow silhouettes;
- **Gait-based pathology classification**: The review of gait-based pathology classification systems includes a discussion on systems that perform classification of gait as either normal or impaired, or across different gait related pathologies. It also includes a review of systems that extract biomechanical gait features with a high level of accuracy.

Apart from the state-of-the-art systems, the review also presents a summary of the publicly available datasets that are used to evaluate the performance of the state-of-the-art systems.

3.2 GAIT-BASED BIOMETRIC RECOGNITION SYSTEMS

The gait-based biometric recognition systems operate such that the gait representations, obtained during feature extraction, best characterise the identity of the individuals being observed. The features obtained from the gait representation can then be used to recognise the observed individual during matching. The quality of the gait representations used for recognition partially depends on the type of acquisition systems used to observe the walking individuals. However, most wearable, floor-based and marker-based acquisition systems that capture the individuals' gait with a high level of accuracy often operate with the cooperation of specialised personnel - see Section 2.3. Thus, this Thesis focuses on 2D vision-based recognition systems that can effectively operate in unconstrained environments.

The state-of-the-art 2D vision-based biometric recognition systems can be broadly classified into two groups, - Figure 3.1:

- **Model-based systems**: They try to model the individuals' body to perform recognition;
- **Appearance-based systems**: They rely on the spatiotemporal information obtained from motion patterns of the individuals to perform recognition.

3.2.1 Model-based systems

Most model-based systems rely on the acquisition of a 3D skeletal model or a complete virtual 3D reconstruction of the individual being observed to perform gait-based biometric recognition.

The 3D skeletal models are typically captured using a depth-sensing camera, as illustrated in Figure 3.2 (a). The static and dynamic features obtained from the 3D model, such as distances between joints and joint angles [68], speed, stride length and variation in barycentre position [69] can then be used to perform recognition, with an accuracy of 98% [68] and 92% [69] over 48 and 52 individuals, respectively. Similar features can also be obtained from a 3D skeletal model constructed using multiple calibrated cameras [70]. However, their performance is inferior to the systems that rely on depth-sensing cameras with a recognition accuracy of only 70%. A more complete/denser 3D model of the individuals can further improve the performance of the 3D model-based recognition systems to 96% over 20 individuals, using a k-NN classifier [71]. The high recognition accuracy is achieved by using affine moment invariant features, obtained from the 2D projections of the 3D model.

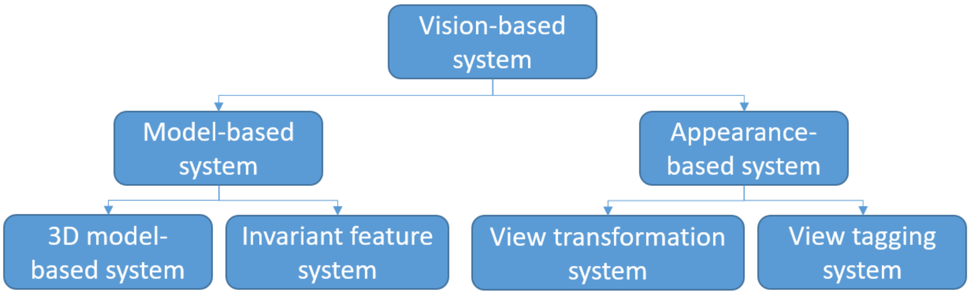


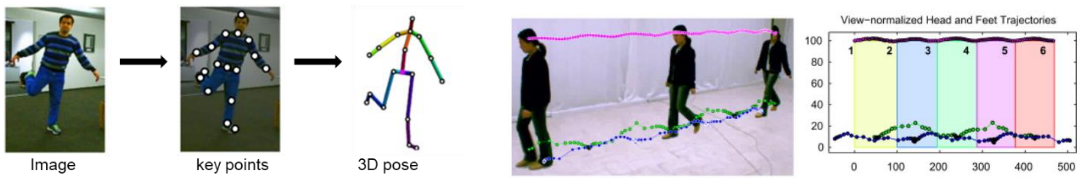
Figure 3.1: Proposed taxonomy for the state-of-the-art vision-based systems.

An advantage of 3D model-based systems is that they can effectively address the problem of viewpoint change using the skeletal models. Since the key points of the skeletal model are estimated by the system in the 3D space, using multiple calibrated 2D cameras [72], or depth-sensing cameras [69], the resulting features are robust to viewpoint changes. Thus, these systems can achieve a high average recognition accuracy of 92% across 5 different views [69]. Similarly, a dense 3D model obtained from multiple 2D cameras can also be effective in tackling viewpoint change [73]. Such systems can transform the features acquired from the model to the desired viewpoint, achieving a recognition accuracy of 75% across 12 different viewpoints of 20 individuals. In situations where only a single 2D camera observes the environment, some 3D model-based systems maintain a database of dense 3D models of the individual captured over time [74]. The features are then obtained with respect to the viewpoint of the camera to perform recognition. Such systems perform well with a recognition accuracy of 97%, when operating across 3 different viewpoints of 42 individuals. These systems can also perform recognition along curved trajectories, which is a significantly more challenging task, with a recognition accuracy of 72%. A second advantage of these systems is that they can also be robust to appearance changes, as the acquisition systems used are usually trained to detect key points over different types of clothing.

A drawback of the 3D model-based systems is that they require calibrations before use, and depth-sensing cameras typically has a limited range of operation, making them ineffective in unconstrained environments. They are also adversely affected by occlusions, as they rely on successful detection of the model key points, which are estimated with respect to one other. The use of 3D models can also result in computationally expensive systems. Finally, the 3D model-based systems also need to consider information, such as camera parameters, distance from the camera, position of the individual, when fitting the model over the observed individuals. Such information is not always available in unconstrained environments.

Other model-based systems, called invariant feature systems, model the gait of the individuals using anatomical positions, such as the head and feet [75] or the hips, knees and ankles [76]. The systems can then obtain features such as angular measurements and movement trajectories of the individuals, to perform gait-based biometric recognition, as illustrated in Figure 3.2 (b). These systems when evaluated over 10, and 20 individuals, result in a recognition accuracy of 98% [75] and 96% [76], respectively. These systems operate well even under viewpoint changes. When individuals are observed from a random viewpoint, the invariant feature systems can model the gait of the individuals in a canonical viewpoint using a rectification step. The resulting features, such as angular measurements and movement trajectories are viewpoint invariant and can be used to perform gait-based biometric recognition successfully. These systems achieve a recognition accuracy of 73.6% across 11 different viewpoints of 124 individuals from CASIA B dataset [77].

However, transformation of gait models into a canonical viewpoint is effective within a small range of viewpoints, which limits their applicability to controlled environments. In addition, the systems that synthesise features for the canonical viewpoint by using a perspective projection model require camera calibrations [78]. Those systems also depend on approximated anatomy ratios to identify the knee and feet positions, which together with the other limitations make them difficult to be used in unconstrained environments.



(a) (b)

Figure 3.2: Gait representations obtained from (a) 3D model-based system, (b) feature invariant systems [75].

3.2.2 Appearance-based systems

Unlike most model-based systems, the appearance-based systems can perform gait-based biometric recognition using the input from a single 2D camera. The spatiotemporal information obtained from the observed individuals, allows these systems to generate gait representations, such as the GEI [79], which can be obtained by averaging cropped binary silhouettes over a gait cycle, as illustrated in Figure 3.3. The use of GEI results in a recognition accuracy of 90% over 124 individuals using a k-NN classifier [80]. Other gait representations, such as chrono gait image (CGI) and frame difference energy image (FDEI), that represent the evolution of the individuals' silhouette over a gait cycle can further improve the recognition accuracy of the appearance-based systems to 94% [80]. Some systems use gait representations, such as the gait entropy image (GEnI) [81], or the multiscale gait image (MGI) [58], which emphasise the dynamic part of the gait by computing the entropy or by applying Gaussian filters to the GEI. Such gait representations significantly improve the performance of the appearance-based systems performance to almost 100%, operating on the CASIA B dataset.

Appearance-based systems can address the problem of change in camera viewpoint by adopting either view transformation or view tagging systems. Systems relying on view transformation, as the name suggests, transform features obtained for a given viewpoint to any other desired viewpoint. The view transformation system applies singular value decomposition to a matrix containing examples of individuals in different views, generating a set of gait feature vectors and a transformation matrix that can transform feature vectors from one viewpoint to another [82].

Recognition in such cases is typically performed by exploring the correlations between and within individuals in different views. The system relies on gait representations such as the GEIs, to obtain the transformation matrix [82]. However, their performance can be improved by using different gait representations, such as the Radon transform-based energy images [83], as well as better performing classifiers, such as support vector machines [84], or multi-layer perceptron [85] achieving a recognition accuracy of 85% [83], 95% [84] and 98% [85], respectively, for 11 different viewpoints of 100 individuals from CASIA B dataset.

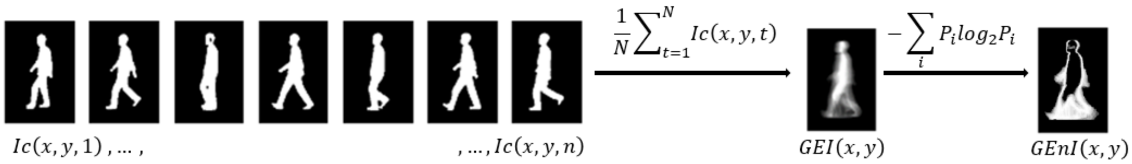


Figure 3.3: Illustration of a GEI and GEnI.

A limitation of view transformation systems is that they tackle the problem of viewpoint change assuming that the viewpoint of the feature to be transformed is already known. This limitation can be dealt with using a view transformation model to generate feature vectors for virtual viewpoints, which are then projected onto a subspace, such as the Grassmann manifold [86], [87], achieving a recognition accuracy of 95% across 11 different viewpoints of 100 individuals from CASIA B dataset. Alternatively, a given silhouette, independently of the view, can be transformed into a canonical viewpoint using a transformation matrix obtained by optimisation techniques, such as the low-rank optimisation of a GTI [88], achieving a recognition accuracy of 85% across 5 viewpoints of 124 individuals from CASIA B dataset. Nonetheless, in the current literature, systems that rely on correlation between views to perform recognition usually cannot address other problems associated with unconstrained environments, such as appearance changes.

View tagging systems, on the other hand, are robust to both changes in the viewpoint of the camera and appearance of the individual being observed, as they follow a two-step approach. The first step of such systems involves viewpoint detection. It is typically performed by analysing the individuals' leg region [89], which is provides a good indicator of the viewpoint, as illustrated in Figure 3.4. Recognition can then be performed with respect to the detected viewpoint of the camera. Thus, the two-step approach restricts comparisons of the individuals' features to the identified view, implying that individuals must be registered in all the considered viewpoints. The use of leg region allows the system to detect 7 different views with an accuracy of 86% and can then perform recognition over 50 individuals with an accuracy of 98%. The performance of the first step of such systems can be further improved by obtaining features that can better represent the viewpoint, such as the entropy of the leg region. It improves the viewpoint detection accuracy to 87% [58] and 89% [90], over 11 different viewpoints from CASIA B dataset. While the performance of the second step, i.e. recognition, can be improved by using better performing classifiers, such as canonical correlation analysis [89] or random subspace learning [58], improving the recognition accuracy to almost 100% over 124 individuals. Recognition can also be improved by using a better gait representation, such as the MGI [58]. Thus, the two-step approach allows the systems to tackle the problem of appearance change in the second step making the view tagging systems suitable for operation in unconstrained environments.

Change in appearance of the individuals can alter their gait representations, as illustrated in Figure 3.5. Most state-of-the-art appearance-based systems that tackle the problem of appearance change do so by training the system using available altered gait representations. The system presented in [91] uses a ratio of amount of body covered by different type of clothing to identify the parts of the individuals' body that will be altered by appearance change. The system can then perform recognition using only the unaltered parts of the individuals' body, achieving an accuracy of 94% and 91% over 124 individuals from CASIA B dataset, carrying a bag and wearing a coat, respectively. However, a limitation of such systems is that they need to be trained with all the available types of clothing that the systems might encounter when deployed. In addition, the appearance of the individuals can be altered by other factors, such as wearing a backpack or carrying a bag, which limits the effectiveness of such systems to clothing-related appearance changes.

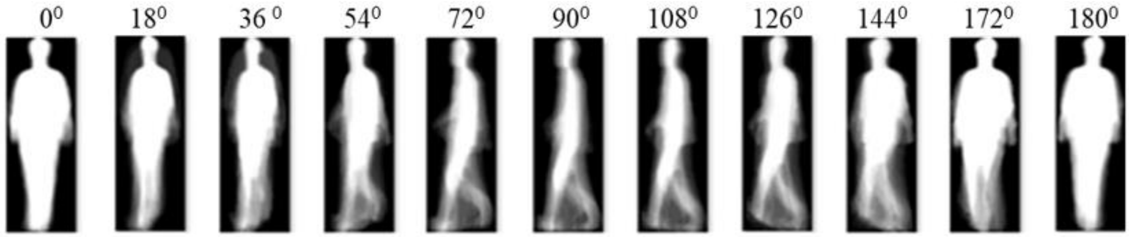


Figure 3.4: GEIs of an individual corresponding to different angles representing the viewpoints of the camera.

To be robust to different types of appearance changes, other appearance-based systems rely on exploring gait representations that highlight the dynamic parts of the individuals' gait, such as the movement of the limbs. Gait representations, such the Poisson random walk [92] and the GEnI [81], [93] can be used in such situations to improve the recognition results. The Poisson random walk performs significantly better than the GEnI, with a recognition accuracy of 93% over 124 individuals carrying a bag. However, they perform poorly when the clothing of the individuals changes, with the recognition accuracy dropping to only 45%. Thus, it can be concluded that these gait representations can tackle the problem of appearance change introduced by static objects such as a backpack but fail when the objects move and alter the appearance of the individuals during walking.


Figure 3.5: GEIs of an individual altered by wearing a coat and carrying a bag.

A significant improvement in recognition results can be obtained by automatically identifying the parts of the individuals that are being altered. Systems, such as [94] can achieve this by automatically setting different weights to different parts of the individuals' body. The system assigns lower weights to the altered parts of the individuals, thus reducing their influence during recognition. This allows the system to achieve a recognition accuracy of 92% and 78% for 124 backpack and coat waring individuals, respectively [94]. The altered parts of the individuals can also be identified by comparing them to the available unaltered sequences [95]. The common parts between the two sequences can then be retained as the unaltered parts of the individuals to perform recognition. The recognition results can also be improved by applying Gaussian filters at different scales [58], which gradually reduces the effect of appearance change by highlighting the unaltered shape of the individuals, resulting in 89% and 76% recognition accuracy for 124 individuals from CASIA B dataset, whose appearance is altered by carrying a bag and wearing a coat, respectively.

The appearance-based systems discussed above are effective in environments where the camera can observe the entire body of the individuals. In situations where the camera is carried by a drone or placed at elevated positions, observing the body of the individuals becomes difficult or even impossible, as illustrated in Figure 3.6. The self-occlusions caused by the individuals can hide the gait features necessary to perform recognition. In such situations, the shadow cast by the individuals can be an alternative source of features for biometric gait recognition. The shadow cast under certain conditions can depict the appearance of the individuals, as discussed in Section 2.4.3. Gait features obtained from such shadows can be successfully used to perform gait-based biometric recognition [22].

Figure 3.6: Image captured using an overhead camera in an unconstrained environment.

The use of the shadows for gait-based biometric recognition involves additional challenges, such as separating individuals from their shadow and dealing with distortions and deformations caused by camera parameters, camera perspective and the time of the day. Using manual segmentation and normalisation of shadow silhouettes and considering a fixed camera viewpoint and a fixed time of the day can lead to good recognition results, but only in constrained environments [96]. The performance of such systems can be improved by introducing automatic shadow segmentation [97] and shadow silhouette normalisation [98]. However, they still require the manual setting of many parameters, such as the position of the sun and the position of the individuals with respect to the camera.

To perform recognition, the shadow silhouette-based biometric recognition systems can use features such as harmonic coefficients obtained from gait stripes [96] or the contour of the shadow silhouette [99]. However, the fusion of the two features provides the best recognition accuracy of 98%, when evaluated with 20 individuals under controlled illumination [100].

To explore further the use of shadows for gait-based biometric recognition, some systems use multiple light sources to cast two shadows, perpendicular to each other [101]. The shadows captured using an overhead camera can then be used to obtain affine moment invariant features, to perform gait-based biometric recognition. Such systems can also tackle the problem of appearance change by assigning different weights to different parts of the shadow based on the influence of their clothing [64].The system presented in [64] achieves a recognition accuracy of 97% over 54 individuals. To address the problem of viewpoint change, some systems use dense 3D models to render the shadow silhouettes of the individuals in the desired viewpoint [102], achieving a recognition accuracy of 99% over 97 individuals. However, the use of additional information, such as the position of the light source and the position of the individuals, still limits their use to constrained environments.

3.2.3 Discussion

From the state-of-the-art review, it can be concluded that the appearance-based systems are best suited to operate in unconstrained environments, as they can operate using a single 2D camera. Since, improving the gait representations improved the performance of most gait-based biometric recognition system, even when using simple classifiers such as k-NN, this Thesis focuses on identifying novel gait representations suitable for appearance-based systems.

This Thesis also focuses on addressing problems such as changes in the viewpoint of the camera and changes in the appearance of the individual, faced by appearance-based systems operating in unconstrained environments. Among the systems discussed in the state-of-the-art review, the view tagging systems are best suited for operation in unconstrained environments. However, the view-detection accuracy of such systems still has a significant margin for improvement. These systems are also limited in their operation as they can detect only the viewpoints registered in the database. In addition, since they depend on the leg region of the individuals, their performance can suffer due to the presence of shadows under the individuals' feet. Thus, the Thesis focuses on improving the view detection step of view tagging systems. It then focuses on automatically identifying the unaltered parts of the individuals' body to tackle the problem of appearance change, as this strategy provides the best results in the state-of-the-art review.

In addition, since the use of shadow silhouettes for gait-based biometric recognition faces several limitations, such as the distortion and deformation of the shadows - see Section 2.4.5, the Thesis focuses on automatically rectifying the shadow silhouettes to a canonical viewpoint before attempting recognition. In addition, since poor segmentation of the individuals from their shadow silhouettes can affect the performance of the recognition systems, the Thesis addresses the problem of shadow segmentation.

3.2.4 Datasets

To evaluate the performance of different gait-based biometric recognition systems, datasets must be collected such that they have enough variation among the captured individuals to obtain representative and statistically reliable results. The literature presents several datasets to perform gait-based biometric recognition, as reported in Table 3.1. However, most of them contain a relatively small number of individuals, such as the CMU MoBo [103], Georgia Tech [104], HID-UMD [105] and CASIA A [106]. One of the first datasets to capture data from more than 100 individuals is the SOTON Large dataset [107]. It captures individuals walking indoor, outdoor and on a treadmill from side and oblique viewpoints. Another dataset containing over 100 individuals, captured in an outdoor environment, is the USF HumanID [108]. The individuals in this dataset are captured from two different viewpoints, walking on concrete and grass surfaces, with and without carrying a bag, and wearing two different types of shoes and captured in 2 different sessions.

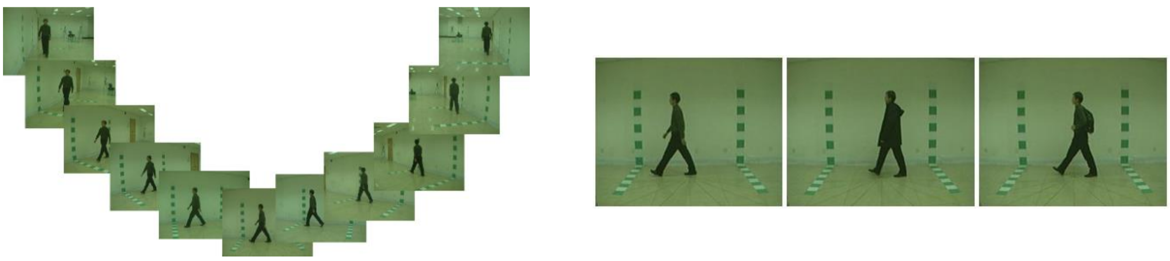
The most popular dataset used to evaluate the gait-based biometric recognition systems is the CASIA B dataset [109], illustrated in Figure 3.7. It contains data from 124 individuals, with 10 walking sequences available for everyone captured during different sessions. These include 2 sequences, where the appearance of the individuals is altered using coats, and in 2 other sequences the appearance of the individuals is altered by carrying backpacks. The dataset also contains a large variation in the observation viewpoints. The individuals are captured from front (0˚) to rear viewpoints (180˚), with an interval of 18˚. Thus, this dataset is suitable for evaluating cross-view gait recognition systems, as well as investigating the effects of appearance changes caused by clothing and carrying items, such as backpacks, on the gait-based biometric recognition systems.

The dataset with the largest number of recorded individuals is the OU-ISIR LP [110] with 4007 individuals captured during 2 different sessions. It contains a wide age range from 1 year old to 94 years old, with an almost balanced gender ratio. The individuals are also captured along 4 different viewpoints (55˚, 65˚, 75˚ and 85˚). Due to the large number of captured individuals, the dataset is suitable to evaluate the recognition accuracy with a high level of statistical reliability.

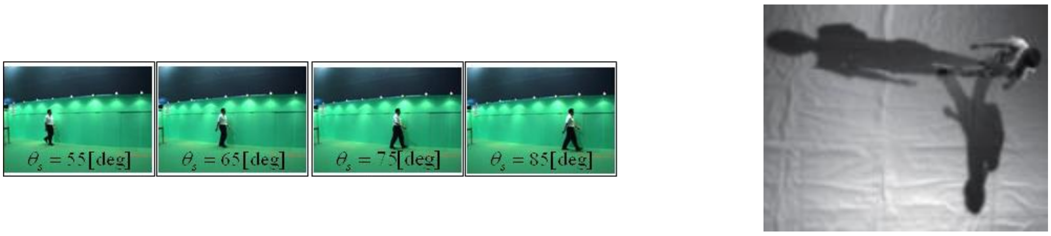
To test the performance of gait-based biometric recognition systems that rely on shadow silhouettes, the literature presents the KY IR shadow dataset [64], which contains shadow silhouettes of 54 different individuals. The dataset is captured from an overhead viewpoint, with a lighting setup such that the individuals cast 2 shadows perpendicular to each other. Everyone in the dataset is captured 6 different times. In the last 2 instances, the appearance of the individual is altered by carrying a bag and changing the clothing respectively. Thus, the dataset can also be used to tackle the problem of appearance change.

Table 3.1: Summary of available datasets for biometric gait recognition

Datasets	**Properties**					
	Year	**Number of individuals**	**Total sequences**	**Number of viewpoints**	**Appearance changes**	**Silhouette type**
CMU MoBo [103]	2001	25	600	6	Y	Body
HID-UMD [105]	2001	25	100	1	N	Body
SOTON Large dataset [107]	2002	115	2128	2	Y	Body
Georgia Tech [104]	2003	15	268	1	Y	Body
CASIA A [106]	2003	20	240	3	N	Body
USF HumanID [108]	2005	122	1870	2	Y	Body
CASIA B dataset [109]	2006	124	1240	11	Y	Body
OU-ISIR LP [110]	2012	4007	7842	4	N	Body
KY IR Shadow dataset [64]	2014	54	324	1	Y	Shadow



(a) (b)



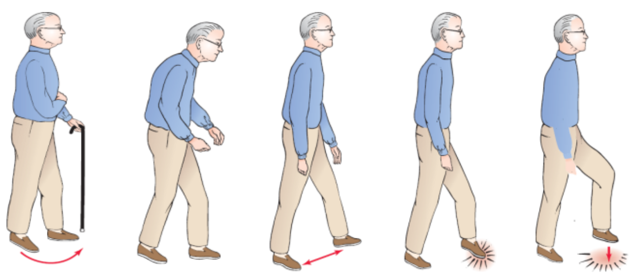
(c) (d)

Figure 3.7: CASIA B dataset representing (a) change in camera's view point and (b) appearance change, (c) OU-ISIR LP dataset representing viewpoint change, (d) KY IR shadow dataset.

3.3 GAIT-BASED PATHOLOGY CLASSIFICATION SYSTEMS

The architecture of gait-based pathology classification systems is similar to the architecture of gait-based biometric recognition systems, illustrated in Figure 2.13. However, unlike recognition systems, pathology classification systems perform classification across different gait-related pathologies in their final matching module, as illustrated in Figure 3.8. A second difference between the two systems lies in the type of features used. While performing recognition, the features acquired by the gait-based biometric recognition systems must best represent the individuals' identity, the pathology classification focuses on acquiring features that best represent the individuals' gait related pathologies.

The gait of the individuals can be acquired by different types of sensors, as discussed in Section 2.3. The wearable and floor-based systems that use accelerometers [27], gyroscopes [41], pressure mats [26] and other sensors can estimate features, such as the duration of the different gait cycle phases, the step count or the step length of the individuals with a high level of accuracy. These features can then be used to classify between different types of walking, such as level, upstairs, and downstairs, walking at different speeds and classify between normal and impaired walking. However, these systems cannot operate effectively in unconstrained environments, as some need the cooperation of clinical personnel to setup the sensors on specific parts of the body, while others need to be setup on the walkways being used. Due to such limitations, this Thesis focuses on vision-based systems to perform gait-based pathology classification.



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

Figure 3.8: Gait representing (a) Hemiplegia, (b) Parkinson's diseases, (c) Cerebellar ataxia, (d) Foot drop and (e) Sensory ataxia [111].

The state-of-the-art review provided in this section discusses the available systems in the current literature, highlighting the features obtained and the type of classifications performed. The review also discusses the systems that focus on the accuracy of the features obtained, when compared to the motion capture system [18]. The motion capture system [18] is considered the gold standard in clinical analysis, due to the accuracy of the features obtained. However, it can effectively operate only in special laboratories, as it requires camera calibrations before use.

Additionally, it is also very sensitive to the lighting conditions. Thus, the review considers other more unconstrained systems classifying them as either model-based or appearance-based systems.

3.3.1 Model-based systems

Model-based systems, similar to those discussed in Section 3.2.1, rely on depth-sensing cameras or multiple calibrated 2D cameras. However, when performing gait-based pathology classification, these systems try to accurately estimate features such as duration of gait cycle, stance phase, swing phase [112], step length, step width and other indicators [113], using the captured 3D skeletal model. These systems can also be used to perform classification of gait as being either normal or impaired using the 3D position of the skeletal joints captured during a gait cycle and achieving a classification accuracy of 98% [47]. Model-based systems that capture gait using multiple 2D calibrated cameras can perform classification of gait related pathologies using features such as the stride length, stride time and stride velocity [114].

Although most model-based pathology classification systems are non-invasive, the limited range of operation (between 80 cm and 4 m) of the depth-sensing cameras and the need for calibration and rectifications of the 2D cameras before use prevents them from being used in unconstrained environments. However, in laboratories, these systems can be used to obtain features, such as knee angles [115], step length and height of the individuals across time [116], with a high level of accuracy. The accuracy of the angles obtained is similar to the motion capture system, which is considered gold standard in clinical analysis [18], within ±3˚ error margin, when evaluated using 20 individuals.

3.3.2 Appearance-based systems

The appearance-based pathology classification systems use spatio-temporal information captured using a single 2D camera to perform classification of gait across different gait related pathologies. The features acquired to perform pathology classification may vary from biometric gait representations, such as GEI [28], to biomechanical features, such as step length, leg angles and gait cycle time [117]. However, biometric gait representations, as the GEI, has been used to classify the observed gait as either Parkinsonian, neuropathic, hemiplegic, diplegic or normal, with a classification accuracy of 74% using SVM classifier [28]. It can also be used to obtain biomechanical features, such as the amount of movement and movement broadness [118], which can then be used to classify gait as being either normal or impaired. In alternative, biomechanical features such as step length, leg angles, can be obtained by fitting a 2D skeletal model using anatomy ratios, onto the silhouette of the walking individuals [54]. The features used can achieve a classification accuracy of 100%, in classifying the gait as either normal or impaired using SVM classifier. Features, such as cadence, speed, and stride length can be obtained within an error margin of 1% by estimating the position of the feet when it is in complete contact with the ground [55]. Classification of gait as either normal or impaired can also be performed with an accuracy of 96% using temporal features, such as the duration of the stance phase and swing phases during a gait cycle [56]. The appearance-based systems can also estimate posture instabilities using features, such as lean and ramp angles [119], axial ratio and change in velocity of the individuals [120], while some other appearance-based systems can distinguish between normal and wavering, faltering or falling gait using features such as homeomorphisms between 2D lattices of binary silhouettes [121].

3.3.3 Discussion

Since the features obtained from a single 2D video camera, contain enough information to characterise individuals' gait, appearance-based systems can be used in unconstrained environments to perform gait-based pathology classification. These systems are much easier to install and operate when compared to the model-based, floor-based or wearable systems. However, most existing appearance-based systems perform only binary classification of gait as either normal or impaired. The features acquired by such systems are also not validated for clinical use. In addition, most systems can only operate on gait captured in a canonical view. Thus, this Thesis focuses on improving the performance of the systems that classify gait across different gait-related pathologies. It also focuses on validating the accuracy of the features obtained for clinical use, while making the system robust to change in camera viewpoints.

3.3.4 Datasets

Since the Thesis focuses on appearance-based systems that perform classification of gait across different gait-related pathologies, it also reviews the available datasets used to evaluate such systems, as reported in Table 3.2.

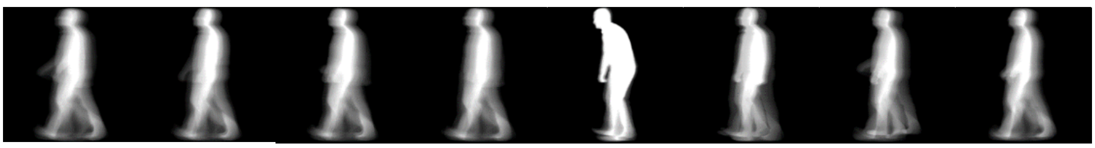
The three reported datasets are captured from a canonical viewpoint. The smallest dataset, the DAI Gait dataset [54], contains 30 gait sequences divided into two groups. The first 15 gait sequences are considered normal, while the remaining 15 contain impaired gait sequences, considering randomly selected pathologies, simulated by 5 walking individuals. The individuals are captured walking over 3 m using the RGB camera of the Kinect sensor.

The next dataset, called the DAI Gait dataset 2 [28], contains healthy subjects simulating gait affected by diplegia, hemiplegia, neuropathy and Parkinson's diseases, together with normal gait sequences, as illustrated in Figure 3.9 (b). Everyone is recorded 3 times walking 8 m, resulting in 75 samples.

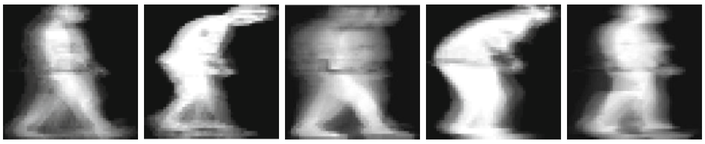
The largest dataset among them is the INIT Gait dataset [118], which contains binary silhouettes of 10 individuals simulating 3 different leg-related gait pathologies, 4 different arm-related pathologies and the normal gait, as illustrated in Figure 3.9 (a). Everyone is recorded 2 different times in a recording studio, at 30 fps, capturing multiple gait cycles in each of these 2 sessions.

Table 3.2: Summary of available datasets for pathology classification

Datasets	Properties				
	Year	**Number of individuals**	**Number of sequences**	**Number of pathologies**	**Total sequences**
DAI [54]	2016	5	3	2	30
DAI 2 [28]	2017	5	3	5	75
INIT [118]	2018	10	2	4	80



(a)



(b)

Figure 3.9: Examples of GEI belonging to (a) DAI 2 and (b) INIT dataset.

Part II: Gait-based biometric recognition

4 GAIT REPRESENTATION PROPOSALS

4.1 INTRODUCTION

The performance of a gait recognition system depends on the quality of gait representations and classifiers used, as discussed in Section 2.5. Most appearance-based systems that provide good recognition results in unconstrained environments use gait representations that capture the evolution of the individuals' gait over a gait cycle, such as the GEI [79]. The recognition results of those systems are further improved by using gait representations, such as the GEnI [81] that highlights the dynamic parts of the individuals' gait, as illustrated in Figure 3.3. Regardless of the gait representation used, most state-of-the-art appearance-based systems rely on a single gait cycle to obtain these representations. Since the gait of an individual when observed across multiple instances can display some variation, the state-of-the-art systems rely on training a classifier to learn such variations, which can then be used to perform recognition.

Two novel proposals of this Thesis explore the possibility of incorporating the variations within and between the individuals into the gait representation.

- **Gait dissimilarity vector (GDV):** It explicitly represents the variation between individuals by exploring the dissimilarity space [122]. It has the additional advantage of hiding other details about the individuals, for instance related to their health condition or gender, which can easily be inferred from the traditionally used gait representations, such as the GEI;
- **Sparse error gait image (SEGI):** It explicitly incorporates variations within an individual's gait across different instances into a single gait representation using RPCA. This can significantly improve the performance of a recognition system, independently of the classifier used.

These novel gait representations are described in the following sections, together with examples of their usage in gait recognition systems.

4.2 GAIT DISSIMILARITY VECTOR

This Thesis presents a new gait representation, called GDV that explores the dissimilarity space. GDV is created using a vector space whose dimensions are defined by pairwise dissimilarity measurements between a given input individual and a set of individuals registered in the so-called representation set, R. Hence, a GDV representation, $D(X, R)$ can be addressed as a data-dependent mapping $D(.,R): X \rightarrow \mathbb{R}^n$ from a given individual, X, to the representation set, obtained using the Euclidean distance. Considering a GEI of the individual as an input, its gait representation can be obtained by computing the pairwise dissimilarities between that input GEI and the GEI of each of the individuals belonging to the representation set, as illustrated in Figure 4.1.

The most important step for successfully obtaining the proposed gait representation involves the selection of the representation set, as it determines the quality of the GDV obtained. The representation set can be selected such that it contains:

- All available GEIs from every registered individual;

- A random set of GEIs;
- A carefully selected set of GEIs, called prototype set.

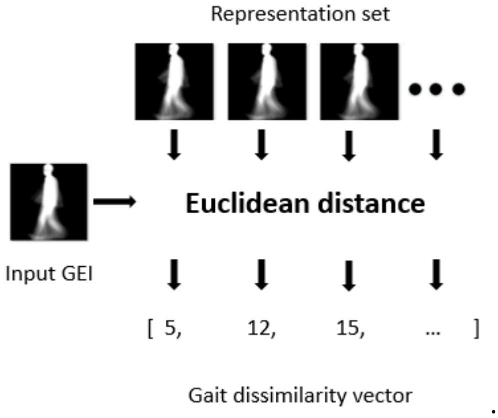


Figure 4.1: Proposed gait dissimilarity vector representation.

The GDV representation proposed in this Thesis follows the third option, selecting the representation set such that it contains one GEI of each registered individual, resulting in a set of prototypes that best represent the variations in the database. Thus, every element of the GDV expresses the degree of difference between the individual being observed and every individual registered in the database. To perform recognition, the proposed GDV representations of the individuals, obtained from their GEIs and the representation set, must be registered into the database during enrolment, as illustrated in Figure 4.2.

Using the GDV gait representation, recognition can then be performed using a classifier trained to distinguish between the registered individuals. To illustrate the usage of the new gait representation, the Thesis considers the use of a k-nearest neighbour (k-NN) classifier, along with principal component analysis (PCA) for dimensionality reduction and linear discriminant analysis (LDA) for data decorrelation, to obtain the recognition results. The use of such a simple classifier enables to highlight the effectiveness of the proposed gait representation.

Recognition is performed by first computing the Euclidean distance between the GEI obtained from the individual to be recognised and the GEIs belonging to the representation set, to obtain the GDV. Then, the GDV of the individual to be recognised and the GDVs of the individuals registered in the database are compared using k-NN, and the one with the smallest Euclidean distance is selected as the recognised identity.

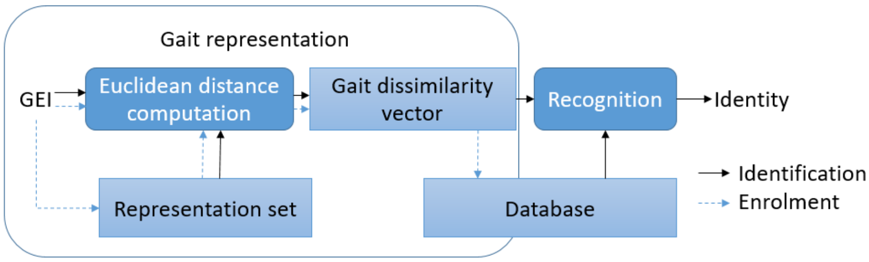


Figure 4.2: Gait recognition using the proposed gait dissimilarity vector (GDV) representation.

4.3 SPARSE ERROR GAIT IMAGE

The second novel gait representation presented in this Thesis, called SEGI, explores the differences between multiple GEIs of the same individual. Gait of an individual, captured across multiple instances, will contain some differences with respect to each other. The differences are significantly larger among different individuals. SEGI represents these differences to perform gait-based biometric recognition.

To obtain the desired differences across multiple instances of the gait of an individual, the proposed SEGI representation uses the GEIs. The differences can then be computed using techniques, such as robust principal component analysis (RPCA) [123]. RPCA creates two types of outputs, a low-rank component and a sparse error component. In the case of SEGI the input to RPCA is a set of GEIs belonging to an individual, as illustrated in Figure 4.3 (a). The low-rank component is generated by projecting the GEIs onto the principal components identified by RPCA. It contains the common parts among all the input GEIs, along with minor variations, such as noise. Thus, it may not contain the most distinguishing features to perform recognition. The sparse error component on the other hand captures the larger differences in the GEIs. These differences cannot be incorporated into the low-rank component, providing a representation of the difference between a specific GEI and the other available GEIs, being called SEGI.

The difference between the GEIs is represented as the variation in the intensity values in the corresponding SEGI representation. A significant difference between GEIs results in strong positive and negative intensity values in the corresponding SEGIs representation - see Figure 4.3 (c). On the other hand, SEGIs obtained from the multiple GEIs belonging to the same individual will typically result in low intensity values, as the difference will not be significant, as illustrated in Figure 4.3 (b). To understand the difference between SEGI representations obtained using multiple GEIs of the same and different individuals, a plot representing the intensity at each pixel value in the SEGI, following a vertical scan from left to right is presented Figure 4.3 (b, c). The middle grey level of the SEGI corresponds to the zero-intensity level in the plot.

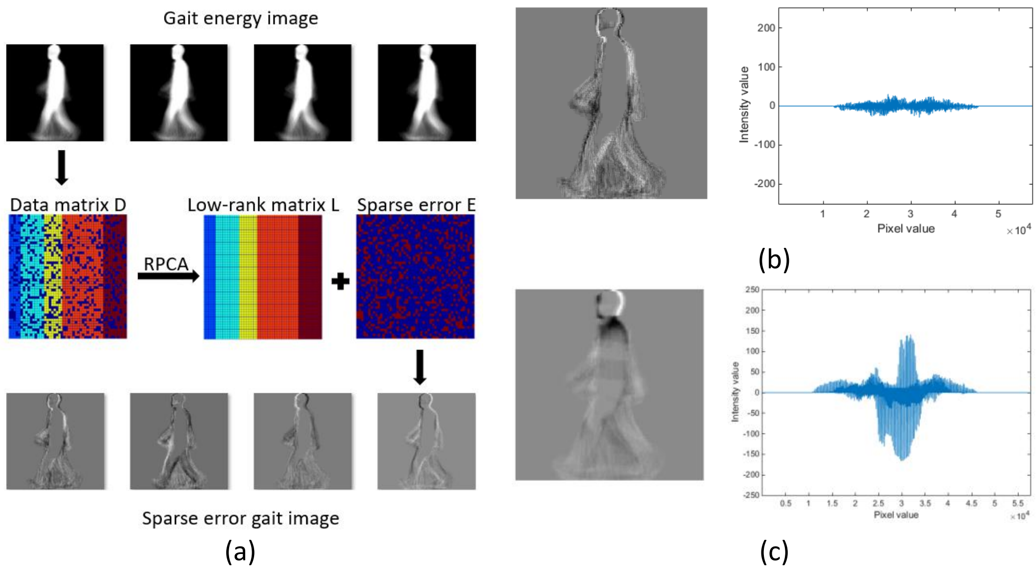


Figure 4.3: (a) Obtaining SEGI by applying RPCA to the GEIs, (b) SEGI representation and their intensity value plot when considering GEIs from the same individual and (c) different individuals.

To use the proposed gait representation for recognition, at least 2 GEIs of everyone are required. The resulting SEGIs can then be used for recognition following the architecture presented in

Figure 4.4. During enrolment, the GEIs of the registered individuals can be transformed into their SEGI representation following the steps described below.

Consider n registered individuals with everyone containing m GEIs represented as 1D-column vectors, obtained by transposing their concatenated rows. They can then be used to construct a matrix D_i for every individual i such that all GEIs I_i belonging to individual i are appended together according to (4.1).

$$D_i = \left[vec\left(I_i^1\right)| \dots |vec(I_i^m)\right] \tag{4.1}$$

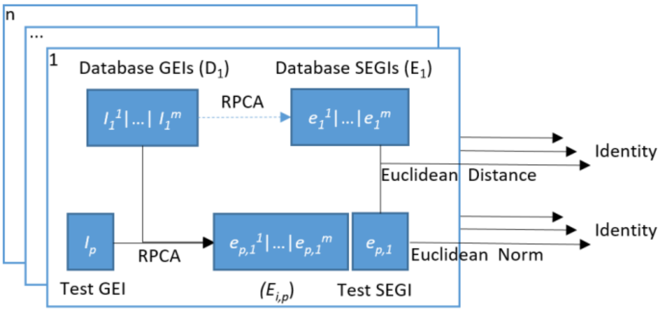


Figure 4.4: Proposed system architecture (SEGI).

The SEGI representation for each individual i can be obtained by applying RPCA to matrix D_i. RPCA decomposes the matrix D_i into $D_i = L_i + E_i$, where L_i is the underlying low-rank matrix and E_i is the sparse error component, following (4.2).

$$\min_{L_i, E_i} rank(L_i) + \gamma \| E_i \|_o \; subject\ to \; D_i = L_i + E_i \tag{4.2}$$

where $\| E_i \|_o$is the counting norm (i.e., the number of non-zero entries in the matrix), $rank(L_i)$ is the number of linearly independent rows or columns in the matrix L_i and γ is a regularisation parameter. However, with the minimisation of rank being a nonconvex problem, the RPCA uses a convex relaxed form, expressed by (4.3)

$$\min_{L_i, E_i} \| L_i \|_* + \lambda \| E_i \|_1 \; subject\ to \; D_i = L_i + E_i \tag{4.3}$$

where $\| L_i \|_*$denotes the nuclear norm of the matrix L_i, i.e., the sum of the singular values of L_i, and $\| E_i \|_1$ denotes the l_1-norm of E_i and λ is a regularisation parameter.

Thus, for each individual i, the use of RPCA results in a matrix E_i that contains the vectorised SEGI representations e_i^j, $j \in 1, \dots, m$, as illustrated in (4.4). The resulting SEGIs can then be registered into the database during enrolment to allow performing recognition.

$$E_i = \left[vec\left(e_i^1\right)| \dots |vec(e_i^m)\right] \tag{4.4}$$

Recognition can be performed over a test GEI I_p, represented as a 1D-column vector, by obtaining its SEGI representation, using the registered GEIs of every individual registered in the database. This is done by appending the test GEI, I_p, to every GEI matrix D_i according to (4.5):

$$D_{i,p} = \left[vec\left(I_i^1\right)| \dots |vec(I_i^m)\middle|vec\left(I_p\right)\right] \tag{4.5}$$

RPCA can then be applied to each matrix $D_{i,p}$ to obtain the corresponding sparse error matrix $E_{i,p}$. The test SEGI, representing the variation between the test GEI I_p and the GEIs belonging to individual i is represented by $e_{p,i}$, according to (4.6):

$$E_{i,p} = \left[vec\left(e_i^1\right)| \dots |vec(e_i^m)\middle|vec\left(e_{p,i}\right)\right] \tag{4.6}$$

When the test GEI belongs to a genuine individual, the resulting SEGI contains low intensity values, as illustrated in Figure 4.3 (c). However, when the test GEI is obtained from an impostor, the resulting SEGI contains high intensity values, due to a significant variation between test and the database GEIs, as illustrated in Figure 4.3 (b).

Recognition can be performed either by using the SEGI intensity values directly or by comparing the test SEGI with the database. The intensity values can be directly used by computing the Euclidean norm for every test SEGI obtained with respect to each individual. The recognition can then be performed by associating the identity of the individual with the smallest Euclidean norm to the SEGI being tested. The second alternative, which typically allows obtaining better recognition results, computes the Euclidean distance between the test SEGI and the database SEGIs with respect to each individual, using a k-NN classifier. Recognition can then be performed by associating the identity of the test GEI to the individual with the smallest Euclidean distance.

4.4 PERFORMANCE EVALUATION

To illustrate the effectiveness of the novel gait representations, they are evaluated using the k-NN classifier and the CASIA B dataset. Even though this dataset contains 11 different viewpoints of 124 individuals, the novel gait representations are evaluated using only the 90˚ viewpoint sequences, as dealing with viewpoint change is a different problem, discussed in Chapter 5. In addition, among the 10 available sequences of each individual, only the first 6 normal sequences are used for the evaluation, as the remaining 4 sequences are altered in appearance by coats and bags, which again constitute a different problem, discussed in Chapter 6. The first 4 sequences of each individual are used for training and the remaining 2 sequences are used for testing, as this is the testing protocol considered by the state-of-the-art systems considered for comparison purposes.

The aim of the evaluation presented below is to understand the advantages of using the GDV and the SEGI representations for gait-based biometric recognition and to compare their performance to that of systems using state-of-the-art gait representations that operate under similar conditions. The results of the evaluation are reported in Table 4.1.

To construct the GDV, a representation set must be selected as discussed in Section 4.2. Since the current dataset has only 124 individuals, a random selection of individuals for the representation set provides poor recognition results. Thus, a representative GEI from each individual is selected for the representation set. To perform the evaluation, the first sequence of each individual is selected to obtain the GEIs for the representation set. The next 3 sequences are used to obtain the GDVs to be registered in the database. The remaining two sequences of each individual are then used to evaluate the recognition performance when using the GDV gait representation.

The SEGI representations for each registered individual are obtained by applying the RPCA to the 4 training GEIs. Recognition can then be performed by either considering the intensity values of the test SEGI using the Euclidean norm or by comparing the training and the test SEGI representations using the Euclidean distance approach. As reported in Table 4.1, the Euclidean distance approach performs better than the Euclidean norm approach, improving the recognition accuracy by 0.8%. However, it should be noted that both approaches perform significantly better than most state-of-the-art gait representations, achieving a recognition accuracy of 98.4% and 99.2%, respectively.

As reported in Table 4.1, the two proposed gait representations perform better than most state-of-the-art alternatives and are equivalent to the MGI [58], which has a recognition accuracy of 100%. However, the main advantage of the proposed gait representations lies in their simplicity. The MGI is obtained by applying Gaussian filters to a GEI at 5 different scales. It then selects 500 random principal components, using 2D PCA, followed by 2D LDA. The selection process is repeated 10 times selecting a new set of 500 random principal components at each iteration. The recognition is then performed by a majoring voting policy. In comparison, the proposed GDV selects a limited number of principal components, that cover 95% of the total variance explained by each principal component, and through a single application of LDA provides a recognition accuracy of 99.6%. A second advantage of the GDV is that it only represents the identity of the individuals. As discussed in Section 3.3.2, the GEI and other similar representations can disclose additional information about the individuals, such as health issues, along with their identity, which most individuals would like to keep private.

The proposed SEGI representation is also computationally inexpensive, with RPCA taking an average 0.03 s to obtain SEGIs using 4 GEIs on an Intel core i7 CPU @ 3.60 GHz, while the average time to perform recognition using the Euclidean norm approach is 9.5 s and the Euclidean distance approach is 9.6 s respectively. The average time to perform recognition using MGI, using the same CASIA B dataset, is reported to be 52.8 s, on Intel core i7 CPU @ 2.93 GHz [58]. However, the average time required just by the WRSL module, when implemented on Intel core i7 CPU @ 3.60 GHz using GEIs as features is 20.8 sec, which is more than double the time required by the SEGI representation.

The proposed SEGI representation is also evaluated to analyse its performance based on the number of available database sequences for each individual. This evaluation is performed by starting with a single database SEGI per individual. The number of database sequences used to compute an individual's SEGI is increased from 1 to a maximum of 5, since only 6 normal sequences are available in the CASIA B dataset. The results are illustrated in Figure 4.5 as a plot of recognition accuracy versus number of database sequences used for SEGI computation per individual. It can be concluded from the plot that using the Euclidean distance provides better results than the Euclidean norm approach. Moreover, if sufficient SEGI representations are made available in the database, a recognition accuracy of 100% can be achieved.

Table 4.1: Recognition accuracy of the proposed and state-of-the-art gait representations.

Recognition systems	**Recognition accuracy** (%)
GDV + KNN (4.2)	80.2
CGI [80]	87.0
GEI [79]	90.0
GEnI [81]	92.3
GPPE [80]	93.3
FDEI [80]	94.0
ARA [124]	94.9
AMI [94]	97.0
SEGI + Euclidean Norm (4.3)	98.4
SEGI + Euclidean distance (4.3)	99.2
GDV + PCA + LDA (4.2)	**99.6**
MGI+WRSL [58]	**100.0**
5 SEGIs + Euclidean distance (4.3)	**100.0**

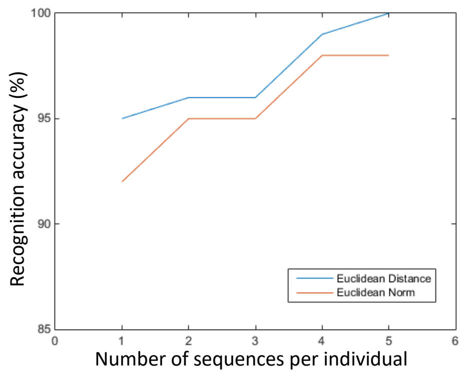


Figure 4.5: Change in recognition accuracy w.r.t. increase in the number of registered sequences per individual.

A drawback of the proposed representations is that they require at least 2 sequences of each individual to be registered in the database. The GDV also requires many enrolled individuals to obtain a good representation set. This prevents the evaluation of these representations using datasets, such as OU-ISIR LP [110], as well as the shadow IST shadow gait dataset, presented in Section 7.4. A second drawback of these representations is their inability to tackle appearance change. Individuals with altered appearance will seldom be recognised using the proposed representations.

However, the proposed GDV and SEGI representation provide a recognition performance improvement over the state-of-the-art gait representations. They are computationally inexpensive and, even using the simple k-NN classifier, achieve recognition accuracies equivalent to the best state-of-the-art gait-based biometric recognition systems, which use computationally expensive and better performing classifiers. The simplicity of the classifier used highlights the quality of the proposed gait representations. The GDV also provides an added advantage of acquiring only the identity information from the individuals’ gait, thus protecting their privacy.

5 PROPOSALS TO TACKLE CHANGES IN THE CAMERA VIEWPOINT

5.1 INTRODUCTION

The performance of gait-based biometric recognition systems can be affected by changes in the viewpoint of the camera observing the individuals. A possible solution to this problem is to use view tagging systems, which perform recognition by first detecting the camera viewpoint. As discussed in Section 3.2, the state-of-the-art systems perform viewpoint detection by using features that represent the registered individuals' leg region [89], [58], [90]. Better performance can be achieved by using features that represent the general structure of the leg region, while also reducing the systems' computational complexity. A limitation with the current state-of-the-art systems is that they are not robust to problems such as different types of clothing, which can cover the legs of the individuals and adversely affect the systems that rely on the leg region. In addition, the presence of shadows under the feet can also hamper performance. Finally, since the systems cannot be trained considering all possible camera viewpoints, a range of viewpoints are usually represented within a single group and insufficient training data may hamper the resulting performance.

To address these problems, the first proposal considers training the system with hash values computed from the observed leg region, which provide a representation of the camera viewpoint. These representations are obtained applying a PHash function over the leg region of the input GEI. However, since the proposed system depends on training for viewpoint detection, a second system is presented in this Thesis that performs viewpoint detection without any training. It observes the position of the individuals' feet over time to estimate the dominant walking trajectory. The direction of the walking trajectory can then be used to detect the viewpoint of the camera. The third system improves on the second one by performing viewpoint detection even in the presence of shadows under the feet. The three systems are presented in the following sections.

5.2 VIEWPOINT DETECTION USING A PERCEPTUAL HASH

The structure of the leg region, as observed in a GEI, provides significant information about the camera's viewpoint, as illustrated in Figure 3.4. To improve the performance of viewpoint detection systems, a way to extract the general structure of the leg region is needed. The proposed system achieves this by using a perceptual hash function – PHash. The PHash is a special type of hashing function that generates outputs that are comparable, as opposed to cryptographic hashing [125]. When applied to two similar images, the resulting Phash values will be similar, and the difference between them will represent a measure of the distance between the two images.

The proposed system selects the bottom 33% [89] of the GEI to represent the leg region, as illustrated in Figure 5.1 (a, b). Next, the discrete cosine transform (DCT) can be applied to the leg region to obtain the DCT coefficients, as illustrated in Figure 5.1 (c), with the DCT coefficients corresponding to the lower frequencies representing the most relevant structure information of the leg region. The proposed system discards the higher frequency coefficients to obtain a compact descriptor, as illustrated in Figure 5.1 (d). It retains only the lower order 22×22 DCT coefficients, as they provided the best viewpoint detection results. The coefficients can then be concatenated, using a raster scan order, to create a unidimensional vector. The hash values, for each element of the PHash vector, can then be computed using (5.1):

$$PHash_c = \begin{cases} 0 \; if \; DCT_c \leq \overline{DCT} \\ 1 \; if \; DCT_c > \overline{DCT} \end{cases} \tag{5.1}$$

where the hash bit $PHash_c$ at index c is determined using the unidimensional vector DCT_c and the mean of the unidimensional vector is $\overline{DCT}$.

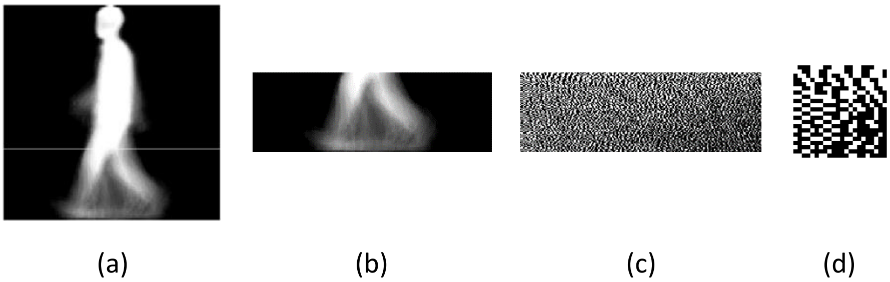


Figure 5.1: Output of the intermediate steps of the PHash function (a) GEI, (b) selected leg region, (c) DCT coefficients, (d) matrix representing the hash bits.

To perform viewpoint detection, the hash values corresponding to each of the available viewpoints are recorded in the database during enrolment, while also training the k-NN classifier, according to the system architecture presented in Figure 5.2. Next, given an input GEI, the proposed system can detect its viewpoint using the k-NN classifier with Hamming distance as the distance measure. The output of viewpoint detection can be used to limit the access of the recognition system to only the relevant part of the database, which is organised with respect to the available viewpoints, thus significantly improving the recognition accuracy of the system.

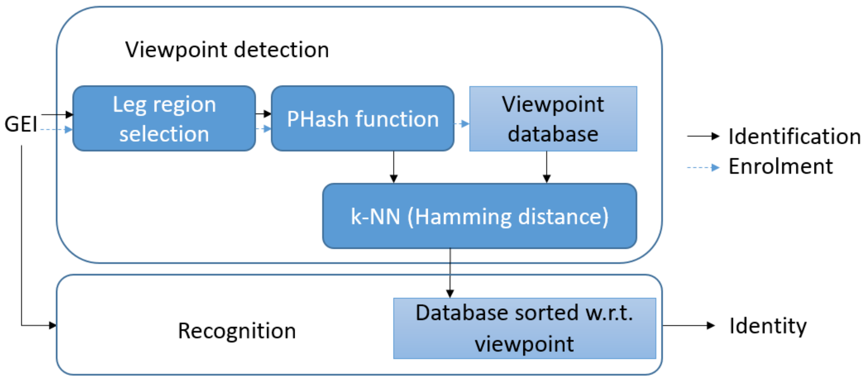


Figure 5.2: Architecture of the proposed viewpoint detection system based on PHash.

5.3 Viewpoint Detection Using Gait Texture Image Contours

Systems that rely on the leg region of the GEI to detect the viewpoint are typically sensitive to certain types of clothing, such as long skirts [89], [58], [90]. Thus, the Thesis presents a novel system that relies on the feet position of the individuals to detect the viewpoint of the camera, according to the system architecture presented in Figure 5.3. The system detects the viewpoint using the direction of the dominant walking trajectory of the individuals, identified using their feet positions. However, if the individual is walking towards or away from the camera the viewpoint is detected by observing the change in the size of the silhouettes over time.

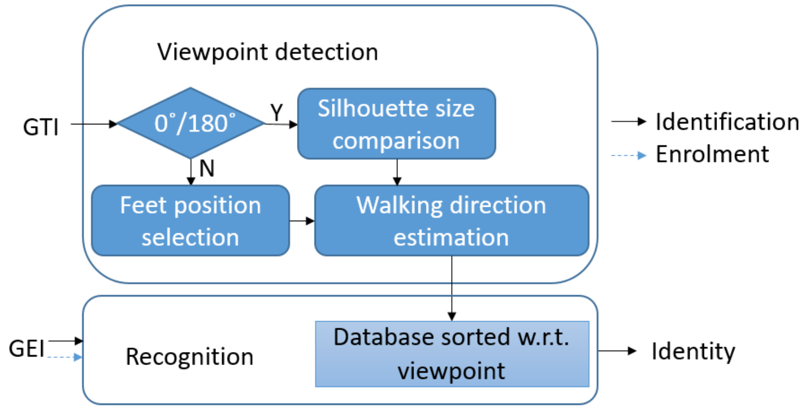


Figure 5.3: Architecture of the proposed viewpoint detection system based on GTI contour.

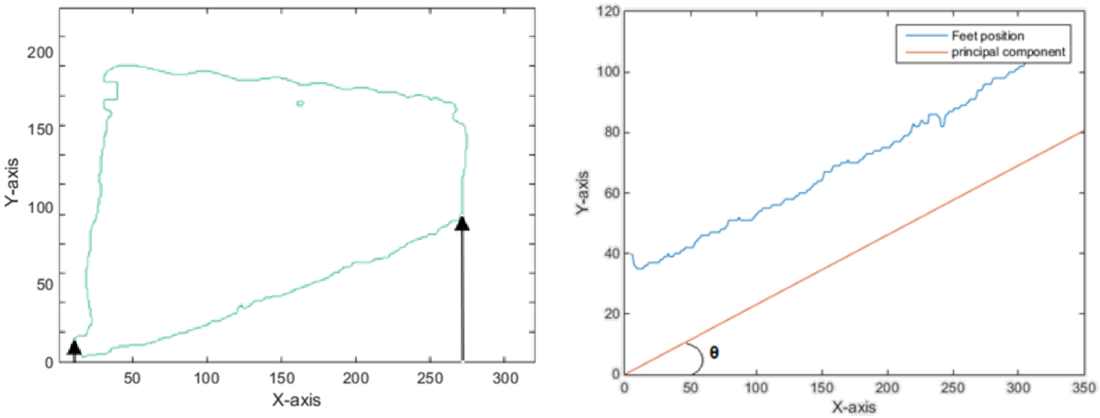
To obtain the evolution of the feet positions over time, the proposed system constructs a gait texture image (GTI) representation. Given T silhouettes $I(x, y, t)$, the $GTI(x, y)$ can be computed by averaging the silhouettes, on their original spatial positions, following (5.2):

$$GTI(x, y) = \frac{1}{T}\sum_{t=1}^{T} I(x, y, t) \tag{5.2}$$

Figure 5.4: Construction of a GTI

The bottom part of the GTI corresponds to the evolution of the feet position of the individuals along time – see Figure 5.4. To select the feet position, the proposed system selects for each column the bottom-most non-zero values of the GTI contour, as illustrated in Figure 5.4 (a). The resulting values can contain positions (notably at its left and right sides) that correspond to the leg or the waist region in the GTI. To filter out these positions, the proposed system thresholds the difference between two consecutive values, discarding the higher difference values as belonging to the leg or the waist regions. The threshold value considered has been empirically set to 10 pixels.

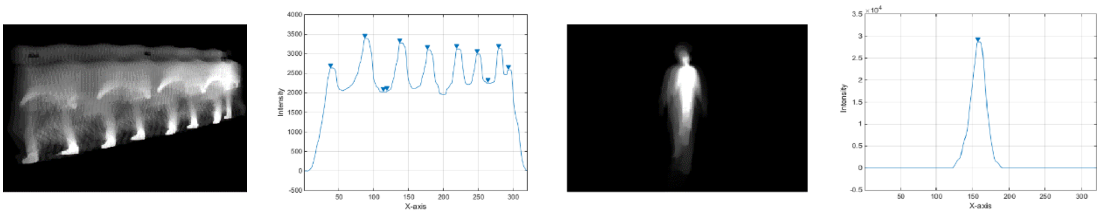
To detect the viewpoint of the camera, the proposed system applies PCA to the filtered feet positions. PCA transforms the feet positions onto a new coordinate space, such that the first principal component captures the largest variance in the feet positions, and therefore its direction θ represents the dominant walking trajectory, as illustrated in Figure 5.5. The viewpoint of the camera can be estimated by associating it to the identified walking direction.



(a) (b)

Figure 5.5: Example of (a) selecting the feet position from the GTI contour and (b) application of PCA to the selected feet positions

The viewpoint can be detected by the proposed system, except when the individuals walk towards or away from the camera, in such situations, the silhouettes will be overlapped in the GTI, not allowing tracing the feet positions, as illustrated Figure 5.6(c). The proposed system detects such situations by checking the number of peaks that appear in the plot of summation of intensities of the GTI along the y-axis. When the individuals walk towards or away from the camera, the proposed system detects a single peak in the intensity plot, while all other viewpoints are represented by multiple peaks, as illustrated in Figure 5.6 (b, d). Finally, the distinction between individuals walking towards or away from the camera can simply be made by comparing the size of the first and last silhouette of the sequence.



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

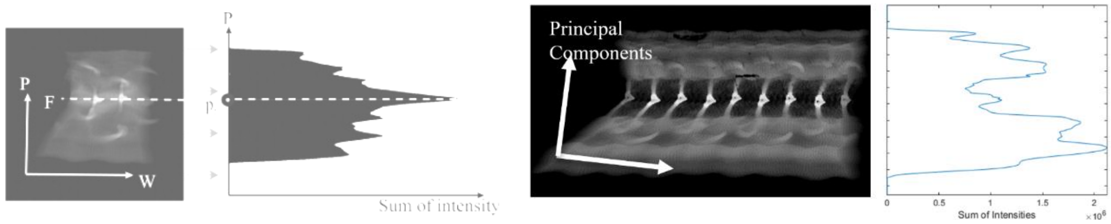
Figure 5.6: GTI representing (a) 126˚ viewpoint, (c) 0˚ viewpoint, and the corresponding vertical sums of intensities for (b) 126˚ viewpoint and (d) 0˚ viewpoint GTI.

Since, the proposed system estimates the viewpoint directly from the first principal component; it does not require any training to perform viewpoint detection. Once, the viewpoint is detected the proposed system can perform recognition with respect to the detected viewpoint. However, the GTI is not descriptive enough to represent the individuals' identity. Thus, the proposed system must use a different gait representation, such as GEI, to perform recognition.

5.4 Viewpoint Detection in the Presence of Shadow

The state-of-the-art viewpoint detection systems that rely on analysis of the leg region perform poorly in the presence of the shadows under the feet [89], [58], [90]. To improve the performance of those systems, shadow segmentation techniques, such as [100], can be used, which performs shadow segmentation by detecting the highest summation value, computed along the first principal component of a GTI as the feet position – see Figure 5.7 (a). As proposed in Section 5.3, the direction of the principal component can be used to detect the viewpoint of the camera, allowing the shadow segmentation technique to perform viewpoint detection.

However, as illustrated in Figure 5.7 (b), the shadow segmentation technique presented in [100] can fail as the principal component computed over the GTI does not always align itself with the walking direction and, when aligned, the highest summed intensity value may not always correspond to the individuals' feet position. Thus, a novel system is proposed in the Thesis that improves on the work presented in [100].



(a) (b)

Figure 5.7: Illustration of (a) success and (b) failure of the system presented in [100]

In Figure 5.7, the positions where the feet of the individuals are in complete contact with the ground are represented by the higher intensity values in the GTI. Therefore, the proposed system uses an initial thresholding operation to identify those feet positions. The threshold is selected as 80% (empirically determined) of the highest intensity value in the image, as it usually corresponds to the feet positions of the individuals in the GTI. The system then fits a line through the centroids of the various feet positions, separating the body (above the line) from the shadow silhouettes (below the line), as illustrated in Figure 5.8. Under certain conditions, some of the high intensity values may correspond to the arm position, due to the overlapping silhouettes in the GTI. Thus, the proposed system uses a random sample consensus (RANSAC) line fitting algorithm [126] to fit the line through the identified feet positions. RANSAC classifies the centroids into inliers and outliers by obtaining a consensus over the fitted line. The resulting line passes through the inliers, while discarding the falsely identified feet positions as outliers.

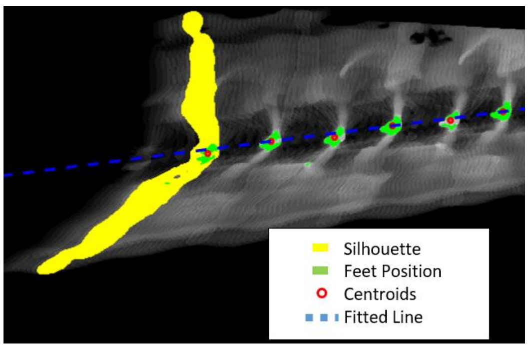


Figure 5.8: Fitting a line through the identified feet positions using RANSAC.

The feet position of the individual can also be determined by analysing the relative orientation of the body and the shadow silhouette. Since the shadow silhouette is projected onto the ground, connecting to the body silhouette at the feet position, the orientation of the body silhouette differs from that of the shadow silhouette projected onto the ground, especially when the legs of the individual are spread wide apart. Thus, the feet position can be identified as the point on the silhouette's contour that exhibits a sudden change in orientation. The proposed system considers only the lower part of the contour, i.e., the bottom-most non-zero values, and selects the contour point with highest y-coordinate value as the feet position, as illustrated in

Figure 5.9 (a). However, during a gait cycle, the arm or knee positions may be falsely identified as the feet position. The system filters out those positions by thresholding the difference between two consecutive y coordinate values with a threshold of 50 pixels, as it provided the best results. A line can then be fit through the feet positions using RANSAC.

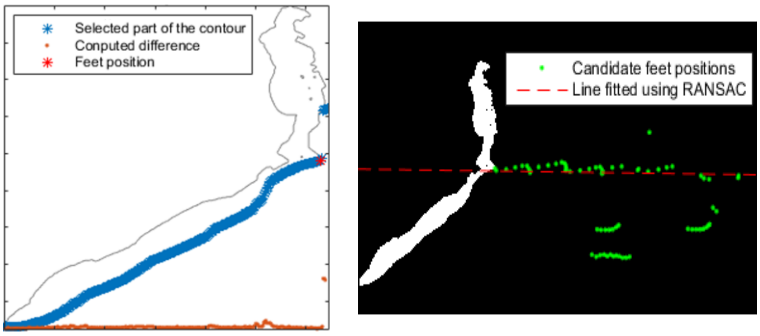


Figure 5.9: Illustration of the intermediate steps of the proposed system depicting (a) feet position detection, (b) line fitting using RANSAC.

The line separating the shadow and body silhouettes represents the dominant walking trajectory of the individuals. Thus, instead of using the proposed system just for shadow segmentation, it can also be used to estimate the viewpoint by computing the direction of fitted line, similarly to Section 5.3. Recognition can then be performed using the segmented silhouettes and the detected viewpoint.

5.5 PERFORMANCE EVALUATION

To evaluate the accuracy of the proposed viewpoint detection systems, CASIA B dataset can be used. The dataset has been captured in a well-illuminated environment where the individuals being observed do not cast a shadow. Thus, only the first two systems proposed in Section 5.2 and 5.3 can be evaluated using the CASIA B dataset. To evaluate the system proposed in Section 5.4, a new dataset, later presented in Section 7.5, can be used.

As discussed in Section 3.2.4, CASIA B dataset contains sequences captured from 11 different viewpoints, of 124 individuals. The system proposed in Section 5.2 uses the first 4 normal gait sequences for training and uses the remaining 6 sequences for testing, following the protocol presented in [58]. Since the viewpoint detection system proposed in Section 5.3, using GTI contour, does not require any training, the entire dataset is used to test the system. The evaluation in this case is performed by quantising the estimated direction of the walking trajectory to the nearest multiple of 18˚, as considered in the ground-truth. The results of the evaluation are reported in Table 5.1 where each system reports correct viewpoint detection accuracy for normal sequences (N) and sequences altered by coats (C) and bags (B) in each column. Table 5.1 also reports the performance of the state-of-the-art systems that operate following the protocol presented in [58].

From the mean results of the normal sequences reported in Table 5.1, it can be concluded that the two proposed systems perform significantly better than the state-of-the-art, with an average detection accuracy of 98% and 97%, respectively. The significant improvement in the performance can be attributed to the use of the PHash function and the detection of the feet position by the proposed systems. The best performance among the state-of-the-art systems is also achieved using normal sequences, with an accuracy of 86% and 89%, respectively. However, their performance degrades due to the presence of bags and coat covering parts of the leg region, as observed for the results included in Table 5.1. The proposed systems address this problem by using the general structure of the leg region acquired by the PHash and by using the trajectory of the feet position, resulting in an average accuracy of 96% and 97% for the bag and coat sequences, respectively.

Table 5.1: Correct viewpoint detection accuracy (%) for N - Normal, C - Coat and B - Bag sequences.

View (˚)	Entropy system [58]			Improved entropy system [90]			PHash system (5.2)			GTI contour system (5.3)		
	N	**C**	**B**	**N**	**C**	**B**	**N**	**C**	**B**	**N**	**C**	**B**
0	83	80	79	89	79	89	99	98	97	98	96	98
18	94	87	85	98	84	92	99	97	96	99	99	99
36	88	85	80	89	79	85	97	96	92	98	98	98
54	92	90	89	98	95	95	98	98	93	99	99	98
72	81	80	78	90	89	88	99	97	98	98	99	98
90	89	79	72	86	69	73	94	92	90	97	97	97
108	79	75	70	85	64	69	98	98	96	95	94	95
126	90	88	85	93	83	89	99	98	98	98	99	99
144	83	81	79	85	73	82	96	95	93	96	96	96
162	89	86	84	88	88	85	97	94	93	97	96	99
180	82	80	75	88	86	76	99	99	98	99	97	99
Mean	86	82	78	89	80	83	**98**	**97**	95	97	**97**	**97**

The state-of-the-art system presented in [89] follows a different protocol to evaluate its performance. It splits the dataset such that 60% of the total available individuals are used for training and the remaining 40% are considered for testing. The system discards some of the viewpoints, operating only on sequences between 36˚ to 144˚, and reports an average accuracy of 86%, 84% and 86% for normal, coat and bag sequences, respectively. Adopting the protocol presented in [89], the proposed PHash system achieves an average accuracy of 95% for normal sequences, 94% for coat sequences and 93% for bag sequences. The proposed GTI contour system performs significantly better with an average accuracy of 97% across all 3 types of sequences.

An additional advantage of the proposed PHash system lies in its low computational complexity. The system is compared to a reimplementation of the entropy system [58] using a computer with an Intel(R) Core(TM) I7 @ 3.60GHz CPU, with 32GB of RAM, in both cases running the same MATLAB R2014b code, just replacing the feature selection by either PHash or the entropy. The proposed PHash system performs the viewpoint detection, with an average time of 0.02s, compared to 0.3s required by the entropy system [58]. Thus, it can be concluded that the proposed PHash system is computationally inexpensive, while performing significantly better than the entropy system [58]. The proposed GTI contour system performs the best with an average detection accuracy of 97%, as reported in the Table 5.1 . However, the construction of the GTI makes the system computationally more expensive, with an average time of 0.4s to perform viewpoint detection.

The Thesis also evaluates the performance of the systems using the OU-ISIR LP dataset [110], which contains 4016 individuals walking along 4 different viewpoints – see Section 3.2.4. However, since the proposed GTI contour system proposed in Section 5.3 does not operate on cropped silhouettes, only the proposed PHash system is evaluated using this dataset. The dataset also presents a protocol to evaluate and compare the performance of the proposed system to the state-of-the-art. According to the protocol, the dataset is divided into 5 predefined subset of 1912 individuals. Each subset is further divided into 2 disjoint groups of 956 individuals each. Following 2 fold cross validation for each subset, the final viewpoint detection accuracy of the system is obtained by averaging the 10 results.

Since the silhouettes of the OU-ISIR LP dataset are rectified to compensate for the viewpoint changes, the shape information in the leg region of the GEI is distorted and thus the bottom third of the GEI (leg region) produces poor results. Therefore, for this dataset, the shape of the entire GEI is considered for computing the hash values, achieving a correct viewpoint detection accuracy of 80%. The resulting confusion matrix is presented in Table 5.2, which reports that most misclassifications occur within the neighbouring viewpoints. The misclassifications typically occur because of the visual similarities between the GEIs belonging to 75˚ and 85˚. The GEIs belonging to the 55˚ viewpoint appear much more distinct, resulting in a viewpoint detection accuracy of 93%, along that view. The visual similarity between the neighbouring viewpoints allows a generic recognition system using a GEI gait representation and a k-NN classifier, to achieve a correct recognition accuracy of 97%, when used with the proposed viewpoint detection system.

Table 5.2: Confusion matrix for the proposed PHash system operating on OU-ISIR LP dataset.

Viewpoints (˚)		**Ground truth**			
		55	**65**	**75**	**85**
Detected	**55**	93	7	0	0
	65	6	83	11	0
	75	0	11	71	18
	85	0	1	26	73

To evaluate the performance of the third proposed view detection system, which is robust to the shadow under the feet, the IST shadow gait dataset, to be presented in Section 7.5 can be used. The dataset contains sequences from 21 individuals, captured from two different viewpoints. The first viewpoint corresponds to the individual walking along a lateral viewpoint of the camera (i.e., 0˚). The second viewpoint is that of the individual walking a diagonal to the camera. Since the system proposed in Section 5.4 does not require training, it performs camera viewpoint detection depending on the angle obtained from the observed direction of walking. If the observed angle is approximately equal to 0˚, the viewpoint is detected as the lateral viewpoint else, it is detected as the diagonal viewpoint. Since the classification is binary and the separation between the two viewpoints is significant, the proposed system performs viewpoint detection with an accuracy of 100%.

To observe the influence of the proposed viewpoint detection system on the recognition results, a generic recognition system [80] that uses a GEI gait representation and a k-NN classifier is evaluated with and without the use of the proposed viewpoint detection system proposed in Section 5.4. The GEIs are composed of body silhouettes obtained by using the line fitted through the feet position separating the individuals from their shadows. For performance evaluation, the system follows the protocol presented in Section 7.5. First, the recognition is performed without viewpoint detection. Next, the system proposed in Section 5.4 is used to identify the viewpoint of the input GEIs. Recognition is performed by limiting the performed comparison to a subset of the database identified by the detected viewpoint. This improves the recognition accuracy of the system as reported in Table 5.3.

Table 5.3 Recognition accuracy of recognition system with and without viewpoint detection

Recognition System	Recognition accuracy
Without viewpoint detection	64%
With viewpoint detection (5.4)	**69%**

It should be noted that in the current evaluation, the dataset considered contains only 2 viewpoints. However, the use of the proposed viewpoint detection system improves the recognition accuracy by 5%. With increase in the number of viewpoints of the individuals registered in the database, more individuals will be falsely matched. Use of the proposed system can provide a gain in the performance under such conditions by identifying the viewpoint of the camera and limiting the matching step to the identified viewpoint.

The three novel systems proposed in this Thesis provide some advantages over each other. The PHash systems proposed in Section 5.2 is the only system among the 3 that can operate on cropped silhouettes, making it is ideal for situations, such as when the individuals walk on a treadmill or the acquisition system provides only cropped silhouettes. The GTI contour systems proposed in 5.3 can operate without any training. Thus, it can detect all possible viewpoints of the camera. The third system proposed in Section 5.4 is the only system equipped to operate in the presence of shadow under the feet. However, the system can operate only for near lateral viewpoints. It cannot detect viewpoint when the individuals walk towards or away from the camera due to a complete overlapping of the silhouettes.

6 PROPOSAL TO TACKLE CHANGE IN THE APPEARANCE OF THE INDIVIDUALS

6.1 INTRODUCTION

The appearance of individuals can be altered due to changes in their clothing or other accessories, such as wearing of a coat or carrying a bag. Such appearance changes can significantly affect the performance of a gait-based biometric recognition system. As discussed in Section 3.2.2, most state-of-the-art systems tackle this problem by learning specific types of clothing models or by using features that highlight the dynamic aspect of the human gait. A drawback of such systems is that they cannot be trained with all possible models of clothing, and/or they cannot correctly represent objects that move and alter the appearance of the individuals while walking. The best results reported in the state-of-the-art are obtained when the altered parts of the individuals' silhouettes are automatically identified. However, even those systems are sensitive to the type of appearance change.

To overcome the limitations discussed above, a novel system robust to appearance change is proposed in this Thesis. The system operates in 4 steps, according to the system architecture presented in Figure 6.1:

- It decomposes a GEI into sections;
- It identifies the altered sections of the GEI by comparison to an average GEI image;
- It matches each unaltered section with the database;
- Recognition is performed using a weighted combination of the result of each section.

The following section explains the details of each step.

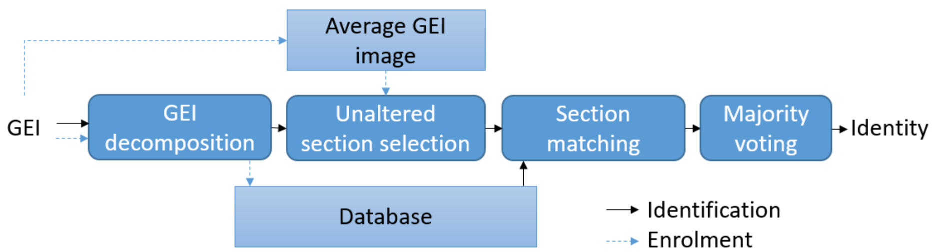


Figure 6.1: Architecture of the proposed gait recognition system robust to appearance changes, based on GEI decomposition

6.2 TACKLING APPEARANCE CHANGES USING GEI DECOMPOSITION

To perform recognition even when the individuals change their appearance the proposed system decomposes the GEI into N horizontal sections of equal size – see. Figure 6.2 (e). The rationale behind splitting the GEI is that appearance change caused by carried bags or wearing coats do not typically alter the entire GEI, but only some small sections of it. Those small sections of the GEI are usually responsible for the poor recognition accuracy of the system, as the altered sections cause the GEIs to be seldom matched to the correct individuals. By decomposing the

GEI into sections, the effect of changes in appearance is limited to only a few of those sections, allowing the rest of the GEI to be successfully matched to the correct individual registered in the database.

Given an input GEI, the proposed system decomposes it into N horizontal sections of equal size, where the number of sections is empirically determined – see Section 6.3. Before matching the sections with the database, the proposed system identifies the altered sections of the GEI using an average GEI image. The average GEI image is obtained by averaging all the GEIs belonging to every individual registered in the database. It represents the general shape of the individuals registered in the database, as illustrated in Figure 6.2 (a). The unaltered sections can then be selected by applying a threshold to the difference between the average GEI image and the input GEI. Since the differences in the altered sections are significantly larger than the unaltered sections, a pixel is considered altered if its difference exceeds the empirically set value of 150. If a section contains altered pixels, that section is discarded. The high threshold value adopted ensures that only the altered sections of the GEI are discarded.

Next, the proposed system applies PCA and LDA to each unaltered section for dimensionality reduction and data decorrelation. The Euclidean distance between each unaltered section and the corresponding sections of the individuals registered in the database are computed, associating the section to the individual with the smallest Euclidean distance.

Finally, recognition is performed by a majority voting decision among the unaltered sections. Each section votes for the individual it is associated with. The input GEI can then be recognised as belonging to the individual with the highest number of votes.

In Figure 6.2 (f), the sections altered by the coat are associated to apparently random individuals, while the unaltered sections corresponding to the leg region of the GEI, are matched with the correct individual (with the identity 1). Thus, in this example the recognition is successfully performed using the majority votes.

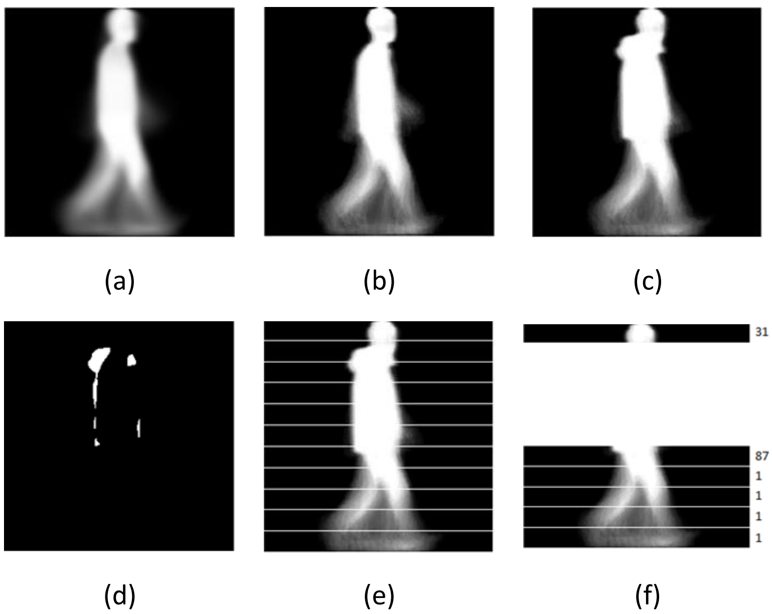


Figure 6.2: Illustration of the operation of the gait recognition system robust to appearance changes based on GEI decomposition: (a) average GEI image, (b) example of an unaltered GEI, (c) example of a GEI altered by wearing a

coat, (d) binary mask representing the altered parts of the GEI, (e) illustration of the GEI sections considered, (f) GEI sections selected for recognition.

6.3 PERFORMANCE EVALUATION

To illustrate the effectiveness of the proposed GEI decomposition system, it is compared to the state-of-the-art systems that are robust to appearance change. Since most of those systems report their performance using only the 90˚ sequences from the CASIA B dataset, the evaluation here is also limited to those sequences. Therefore, the recognition accuracy of the proposed system is computed using the first 4 normal sequences of each individual for training and the remaining 6 sequences for testing.

Before evaluating the system for its robustness to appearance change, a preliminary evaluation can be performed to identify the number of ideal sections, N, into which the GEI should be decomposed. Evaluation is conducted by varying the number of sections from 5 to 20, to obtain that optimal number. The results presented in Figure 6.3 suggest that the ideal number of sections to consider for sequences altered by wearing coat and carrying bags is 13 and 11, respectively. While the performance of the system for normal sequences is indifferent to the number of sections, because all the sections match with correct individual as none of them underwent appearance changes. The default value for the number of sections is set to 11, since it provides the best mean performance across all 90° sequences.

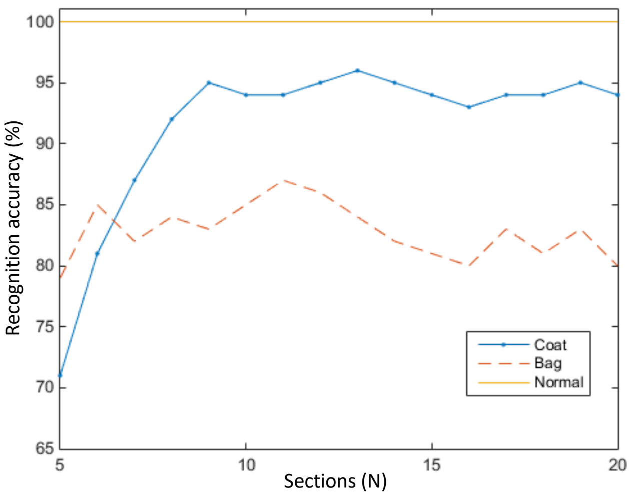


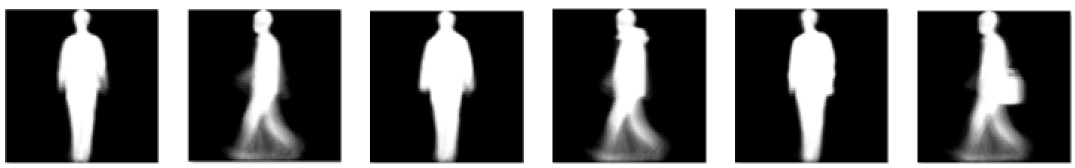
Figure 6.3: Performance of the proposed system with respect to the number of sections N.

The recognition accuracy, using the 90° sequences of the CASIA B dataset, for the proposed and state-of-the-art systems is reported in Table 6.1. The results indicate that the mean recognition accuracy of the proposed GEI decomposition system is the best among the reported state-of-the-art gait recognition systems, which claim to be robust to appearance changes. Most of these systems are effective when performing recognition over normal waking sequences, with an accuracy of almost 100%. The challenge arises in face of appearance changes, where the proposed system performs significantly better than the state-of-the-art, operating on sequences altered by wearing coats, while its performance when carrying bags is also among the best ones listed. The decomposition of the GEI into N sections allows isolation of the sections altered by appearance change, thus improving the performance of the system. Even if the voting is performed without discarding any sections, the proposed system obtains a competitive performance, with a recognition accuracy of 94% and 86% for coats and bags, respectively. It should also be noted that the proposed system is indifferent to the reason for appearance change, while some state-of-the-art systems, such as the MGI [58], use the information about the reason for appearance change to adapt to that situation.

Table 6.1: Recognition accuracy (%) of the proposed and the state-of-the-art systems robust to appearance change.

Recognition System	Recognition accuracy (%)			
	Normal	Coat	Bag	Mean
P. Rw. GEI [92]	-	-	-	93.0
Pal Entropy [93]	93.0	22.0	56.0	57.0
Masked GEIs [95]	100.0	55.0	79.0	78.0
AMI [94]	97.0	78.0	**91.0**	88.0
MGI [58]	100.0	76.0	89.0	88.0
GEI decomposition system (6.2)	**100.0**	**96.0**	87.0	**94.0**

To evaluate further the performance of the proposed system, its recognition accuracy is obtained with respect to every viewpoint available in CASIA B dataset. The evaluation protocol includes using the first 4 normal sequences of every individual for training and the remaining 6 sequences for testing, for all 11 viewpoints of the 124 individuals. As reported in Table 6.2, the proposed system performs well with an average recognition accuracy of almost 90% for all available viewpoints, suggesting that the proposed system can be successfully used along a view tagging system. However, as expected the best results are obtained for 90° sequences with an average recognition accuracy of 94%. The reason for the best results corresponding to the 90° sequences can be associated to an easier detection of the altered sections of the GEI for near lateral viewpoints, as illustrated in Figure 6.4. Also, although the appearance changes caused by carrying bags is easy to detect, as it is more localised, the weight of the bags can additionally cause posture changes, which can adversely affect gait-based recognition systems – see Figure 6.4 (a, b, e and f). However, even under such conditions, the proposed system achieves an average recognition accuracy of 87%.



(a) (b) (c) (d) (e) (f)

Figure 6.4: Lateral and frontal examples of GEIs obtained from (a, b) normal, (c, d) coat and (e, f) bag sequences.

Table 6.2: Recognition accuracy (%) of the proposed system w.r.t. different viewpoints of the camera

Viewpoint (˚)	Recognition accuracy of proposed system (%)			
	Normal	**Coat**	**Bag**	**Mean**
0	100.0	83.0	89.0	**91.0**
18	99.0	84.0	87.0	**90.0**
36	99.0	86.0	86.0	**90.0**
54	99.0	88.0	88.0	**92.0**
72	100.0	92.0	86.0	**93.0**
90	100.0	94.0	87.0	**94.0**
108	100.0	90.0	88.0	**93.0**
126	100.0	87.0	91.0	**93.0**
144	99.0	90.0	88.0	**92.0**
162	100.0	82.0	86.0	**89.0**
180	100.0	88.0	81.0	**90.0**
Mean	**100.0**	**88.0**	**87.0**	

7 PROPOSALS FOR RECTIFICATION OF SHADOW SILHOUETTES

7.1 INTRODUCTION

The use of shadow silhouettes can be an alternative gait representation source, in situations where the body of the individuals is occluded due to factors such as the viewpoint of the camera, or the background properties, as discussed in Section 2.4. The use of shadow silhouettes can also provide an advantage in situations where the individuals are observed over long trajectories. Under such conditions, body silhouettes obtained at two different points can appear significantly different due to the change in the viewpoint of the camera. However, under the same conditions, the features acquired from the shadow silhouettes often appear similar to each other, as illustrated in Figure 7.1 (where the face is hidden to maintain privacy). When observed from a side view position, both the body and the shadow silhouettes can be used in a multimodal system, to improve recognition results.

Figure 7.1: Change in viewpoint of the camera caused by observing individuals walking long distances.

The literature presents several systems that use shadow silhouettes to perform gait-based biometric recognition – see Section 3.2.2. However, those systems are effective only in constrained environments, unable to tackle problems such as changes in orientation, skewness, scale and appearance of the shadow silhouettes caused by camera perspective and parameters and change in the position of the individuals with respect to the light source. Thus, the Thesis presents two novel systems to rectify the silhouettes, compensating for the distortions and deformations caused by the camera perspective and parameters. The proposed systems address the problems caused by changes in the position of the light source with respect to the individuals, using the view tagging system presented in Section 5.4, where a database containing shadow silhouettes sorted with respect to the light source position is available.

7.2 RECTIFICATION OF SILHOUETTES USING ROBUST PRINCIPAL COMPONENT

Images acquired by a camera can contain distortions and deformations caused by the camera perspective and parameters, which also affect the shadow silhouettes captured in those images – see Section 2.4.2. The state-of-the-art gait recognition systems operating on shadow silhouettes presented in Section 3.2.2 perform recognition if the training and test sequences contain the same distortions and deformations, which is not always the case in unconstrained environments. Thus, to rectify the silhouettes before performing recognition, the proposed system transforms them into a common canonical view, using a homographic projective transformation. To obtain the transformation parameters directly from the image, the proposed system uses the transform invariant low-rank textures (TILT) [127]. However, TILT only operates on images with regular symmetric patterns. Since a GTI contains regular symmetric patterns, as the first half of the gait cycle roughly repeats in a symmetric way in the second half of the gait cycle, this type representation can be used as input to TILT. The proposed system therefore obtains a gait texture image, $GTI_s(x, y)$, by vertically flipping the K binary images containing the segmented shadow silhouettes, $I_s(x, y, t)$ and averaging them, according to equation (7.1) – see Figure 7.2 (a, b).

$$GTI_s(x, y) = \frac{1}{K}\sum_{t=1}^{K} flip(I_s(x, y, t)) \tag{7.1}$$

The proposed system can then use TILT to obtain a transformation matrix τ from GTI_s by decomposing it into a low rank component GTI_s^o and a sparse error component E, according to (7.2).

$$GTI_s = (GTI_s^o + E).\tau^{-1} \tag{7.2}$$

To obtain the transformation matrix τ, (7.2) can be presented as an optimisation problem (7.3).

$$\min_{GTI_s^o, E, \tau} rank\,(GTI_s^o) + \gamma\|E\|_0\,, \;\; GTI_s.\tau = GTI_s^o + E \tag{7.3}$$

where $\|E\|_0$ denotes the number of nonzero entries in E. The regularisation parameter γ is a trade-off between the rank of the GTI_s^o and the sparsity of error E.

TILT performs optimisation such that it obtains the lowest possible rank for GTI_s^o, while having the fewest possible nonzero entries in E, that agree with the observed GTI_s up to the domain transformation τ. Since the optimisation problem presented in (7.3) is nonconvex, TILT uses the convex relaxed form expressed in (7.4) and solves it via successive convex programming, as detailed in [127].

$$\min_{GTI_s^o, E, \tau} \|GTI_s^o\|_* + \lambda\|E\|_1\,, \;\; GTI_s.\tau = GTI_s^o + E \tag{7.4}$$

where $\|GTI_s^o\|_*$ denotes the nuclear norm, and $\|E\|_1$denotes the l_1 norm.

The resulting transformation matrix τ can then be used to rectify the shadow silhouettes into a canonical view, as illustrated in Figure 7.2 (c, d). The transformation matrix corrects the orientation, rectifies the distortions and deformations caused by camera perspective and parameters.

The rectification of shadow silhouettes into a canonical view has some limitations. Notably, it cannot recover the self-occluded parts of the silhouette, with only the visible part being transformed into its canonical view. Thus, to minimise the impact of having incomplete shadow silhouettes, the proposed system performs recognition using GEIs obtained from the shadow silhouettes. The GEI compensates the missing parts of a silhouette by also considering the other silhouettes within a gait cycle where those parts may be visible. Given an input GEI, recognition can be performed by computing the Euclidean distance between the test GEI and the GEIs registered in the database and associating the identity to the individual with the smallest Euclidean distance. The performance can be further improved by using PCA for dimensionality reduction and LDA for data decorrelation.

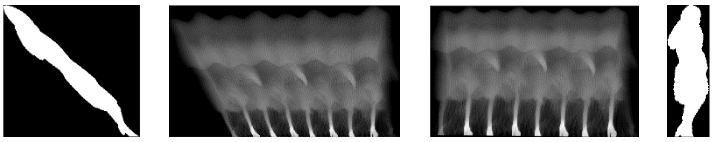


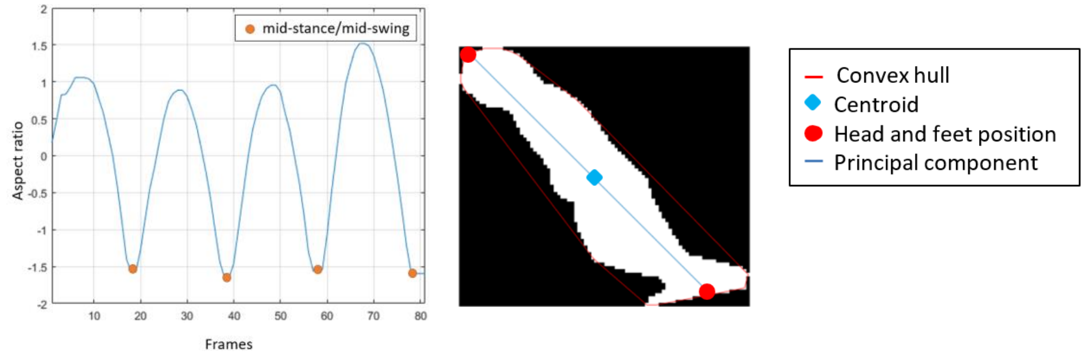
Figure 7.2: Proposed Shadow Silhouette rectification system: (a) flipped shadow silhouette, (b) GTI_s, (c) GTI_s^o, (d) rectified shadow silhouette.

7.3 RECTIFICATION OF SILHOUETTES USING 4-POINT CORRESPONDENCE

Since the approach based on TILT relies on solving an optimisation problem, it cannot be fully controlled. Thus, in the Thesis another novel system that estimates the transformation parameters is proposed, using the head and feet positions in the shadow silhouettes detected at the beginning and at the end of a gait sequence, or portion of a sequence, where the individual is moving in a straight line. To obtain the best estimates for the head and feet positions, the proposed system selects the positions during the mid-stance and the mid-swing phases of the gait cycle, when the arms and feet of the individuals are closer to their body. The proposed system selects shadow silhouettes from the first mid-stance and last mid-swing phases available. It determines the mid-stance and the mid-swing phases by analysing the aspect ratio of the cropped shadow silhouettes along the gait sequence. The aspect ratio in this case is normalised by subtracting the mean and dividing by the standard deviation. The mid-stance/mid-swing correspond to the lowest values within the resulting normalised aspect ratio plot – see Figure 7.3 (a).

Once both the first and the last shadow silhouettes in the mid-stance/mid-swing phases are identified, the proposed system can apply PCA to estimate the orientation, convex hulls and centroids of the selected shadow silhouettes. By fitting a line through the centroid, along the first principal component's direction intersecting the convex hull, the head and feet positions of the respective shadow silhouettes can be estimated, as illustrated in Figure 7.3 (b). Thus, the system obtains the 4-points corresponding to the head and feet position of the shadow silhouettes belonging to the first mid-stance and the last mid-swing phase.

To rectify the silhouettes, the proposed system estimates the location of the 4 points in a canonical view according to (7.5), which allows the system to establish a one-to-one correspondence between the observed and the canonical views, as illustrated in Figure 7.4.



(a) (b)

Figure 7.3: (a) mid-stance/mid-swing detection, (b) head and feet positions selection using the proposed system.

$$
\begin{gathered}
x_l = x_1, \\
x_r = x_1 + \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \\
y_t = (y_1 + y_2)/2\,, \\
y_b = (y_3 + y_4)/2.
\end{gathered}
\tag{7.5}
$$

where the points representing the head and the feet positions of the individual at the first mid-stance phase are (x_1, y_1), (x_3, y_3) and at the last mid-stance phase are (x_2, y_2), (x_4, y_4), respectively. Their mappings onto a canonical view are $(x_l, y_t), (x_l, y_b), (x_r, y_t)$ and (x_r, y_b), respectively.

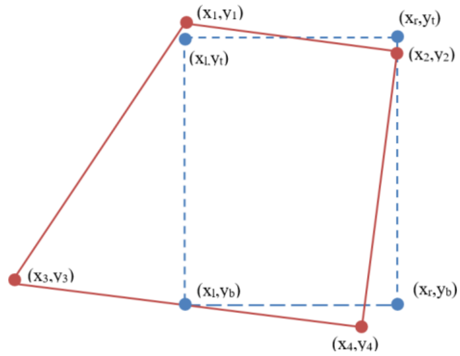


Figure 7.4: The head and feet positions in the observed view (red) and their corresponding mapping in the canonical view (blue).

Using the estimated head and feet positions and their mappings into the canonical view, the proposed system can estimate the parameters for the homographic projective transformation as a matrix H, consisting of three sub-matrices T, R and P [60], where T represents translation, R represents rotation, scaling and shear, and P represents perspective transformations, according to (7.6).

$$
H = \begin{bmatrix} R & P \\ T & 1 \end{bmatrix}
$$

$$
\text{where } R = \begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix},\ T = \begin{bmatrix} t_x & t_y \end{bmatrix},\ P = \begin{bmatrix} p_x \\ p_y \end{bmatrix} \tag{7.6}
$$

The parameters of T, R and P are obtained from (7.7), assuming (x_l, y_t) to be the coordinate axis origin.

$$P = \begin{bmatrix} \dfrac{1}{x_r - 1} \dfrac{\begin{vmatrix} x_1 - x_2 + x_4 - x_3 & x_3 - x_4 \\ y_1 - y_2 + y_4 - y_3 & y_3 - y_4 \end{vmatrix}}{\begin{vmatrix} x_2 - x_4 & x_3 - x_4 \\ y_2 - y_4 & y_3 - y_4 \end{vmatrix}} \\ \dfrac{1}{y_b - 1} \dfrac{\begin{vmatrix} x_2 - x_4 & x_1 - x_2 + x_4 - x_3 \\ y_2 - y_4 & y_1 - y_2 + y_4 - y_3 \end{vmatrix}}{\begin{vmatrix} x_2 - x_4 & x_3 - x_4 \\ y_2 - y_4 & y_3 - y_4 \end{vmatrix}} \end{bmatrix}$$

$$T = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \tag{7.7}$$

$$R = \begin{bmatrix} \dfrac{x_2 - x_1}{x_r - 1} + p_x x_2 & \dfrac{x_3 - x_1}{y_b - 1} + p_y x_3 \\ \dfrac{y_2 - y_1}{x_r - 1} + p_x & \dfrac{y_3 - y_1}{y_b - 1} + p_y \end{bmatrix}$$

Using the estimated parameters, a given point (x_{iv}, y_{iv}) belonging to a shadow silhouette in the observed view can be transformed into a point (x_{cv}, y_{cv})in the canonical view, according to (7.8).

$$x_{cv} = \frac{r_{00} x_{iv} + r_{10} y_{iv} + t_x}{p_x x_{iv} + p_y y_{iv} + 1}$$

$$y_{cv} = \frac{r_{01} x_{iv} + r_{11} y_{iv} + t_y}{p_x x_{iv} + p_y y_{iv} + 1} \tag{7.8}$$

The proposed system, thus, transforms the available shadow silhouettes into a canonical view, correcting for distortions and deformations, such as skewness, scale, orientation. It then allows the system to perform recognition using the shadow silhouettes captured at any part of the gait sequence, which otherwise could be a challenging task. The proposed system can then perform recognition following the steps presented in Section 7.2.

7.4 SHADOW TYPE IDENTIFICATION

As discussed in Section 2.4.3, depending on the light source, individuals can cast either a sharp or a diffused shadow, in an unconstrained environment. The systems proposed in Sections 7.2 and 7.3 can operate only on sharp shadows, which appear similar to the body silhouettes. In addition, the proposed viewpoint detection system, presented in Section 5.4, operates effectively only in the presence of sharp shadows. Thus, there is a need to identify the type of shadow cast, as the presence of diffused shadow surrounding the individuals' feet can hamper the systems' performance. Diffused shadows do not provide useful information but, if identified correctly they can be easily segmented out using systems such as the one proposed in [53]. Thus, this Thesis presents a novel system that identifies the type of shadow cast by the individuals.

The proposed system uses the silhouette, a foreground image (FGI) containing the individual being observed, and a background image (BGI), used as a reference for background subtraction, to identify the type of shadow cast by the individual – see Figure 7.5. It uses the luminance components of the FGI, FGI_i^{ic}, and the BGI, $BGI\ _i^{ic}$ to obtain an intensity ratio μ_i for every pixel value i, according to (7.9).

(a) (b) (c)

Figure 7.5: Illustration of (a) silhouette, (b) background and (c) foreground image.

$$\mu_i = \frac{FGI_i^{ic}}{BGI\ _i^{ic}} \tag{7.9}$$

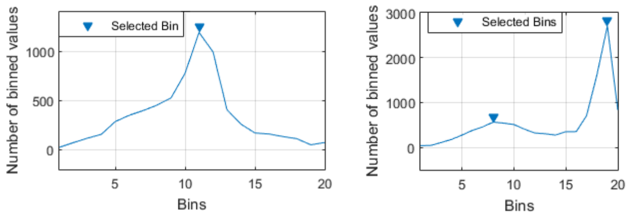
The system then filters the intensity ratio values μ_i using the binary silhouette such that only those values that belong to either the individuals' body or their shadow are retained. A histogram can then be generated for the selected intensity ratio values. The plot, depending on the type of shadow cast by the individual, can contain one of the two distinct shapes, as illustrated in Figure 7.6 (a, c). The difference in the plot shape is caused by the distribution of intensity ratio values in the sharp and the diffused shadow areas. The individual's body area, usually having a significant contrast with the background, as illustrated in Figure 7.5 (b, c), results in low intensity ratio values – see Figure 7.6 (b). On the other hand, a sharp shadow contains a significant contrast with its background and, being a darker area, also changes the background chromaticity resulting in a large number of low intensity ratio values in the histogram plot, as illustrated in Figure 7.6 (a, b).

In the case of a diffused shadow, a significant contrast still exists between the individual's body and the background. However, the diffused shadow only slightly reduces the intensity of the background, not causing significant changes to its chromaticity. Thus, the diffused shadow generates many high intensity ratio values. Therefore, the resulting histogram plot contains a significant amount of both low and high intensity ratio values, as illustrated in Figure 7.6 (c, d).

To identify automatically the type of shadow, the histogram is organised into 20 bins, followed by the application of a Gaussian filter. The bins corresponding to peaks in the histogram are then identified, as illustrated in Figure 7.7. It is observed that the sharp shadows, i.e., those with low intensity ratio values, are represented by histograms with a single peak – see Figure 7.7 (a), while diffused shadows, i.e., containing large numbers of both low and high intensity ratio values, are represented by two clearly separated peaks – see Figure 7.7 (b). Therefore, depending on the number of peaks identified it is possible to have a classification of the type of shadow cast by the walking individual.

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

Figure 7.6: Intensity ratio histogram plot and the corresponding images for (a, b) sharp and (c, d) diffused shadows



(a) (b)

Figure 7.7: Illustration of the two different type of plots obtained for (a) sharp, (b) diffused shadows.

7.5 THE IST SHADOW DATASETS

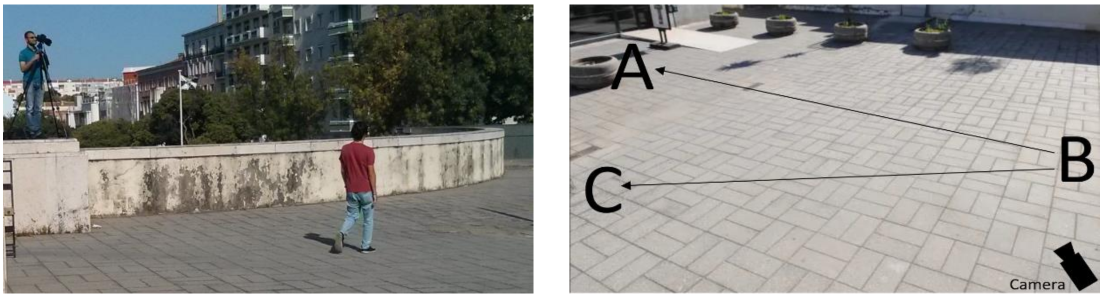
Two new datasets of gait including shadows have been acquired to allow a more complete evaluation of the novel systems proposed in the Thesis. The “IST shadow gait dataset” and the “IST shadow type dataset” are presented in the following.

7.5.1 The IST shadow gait dataset

The proposed systems presented in Sections 7.2, 7.3 and 7.4 address several limitations of the state-of-the-art, such as robustness against viewpoint changes, rectification of shadow silhouettes and shadow type detection. The proposals open the possibility to explore the advantages of using rectified shadow silhouettes as an alternative, or as complement, to using body silhouettes. To test those possibilities, the existing KY IR Shadow dataset [64] can be used. However, it captures gait of the individuals along a single viewpoint for approximately 1 gait cycle. It also lacks an unoccluded view of the body silhouettes, preventing comparisons between systems that use body and shadow silhouettes Thus, to perform a complete evaluation of the proposed systems, a new dataset has been acquired containing unoccluded body and shadow silhouettes, captured along different viewpoints.

The new dataset, called “IST shadow gait dataset”, includes silhouettes captured from 21 individuals. The dataset is collected outdoors, with the setup illustrated in Figure 7.8 (a), where each individual walks along two directions: BA and BC, as illustrated in Figure 7.8 (b), registering sequences captured from two different viewpoints. Video acquisition is performed using a Nikon D300S camera, at 24 fps with a spatial resolution of 640×424 pixels. The data has been collected in July 2017, between 5:30 and 6:30 pm, in the campus of Instituto Superior Técnico, Lisbon, Portugal. Each individual is recorded on 2 different days (sessions). Everyone walks 3 times in each direction during each session, amounting to 12 gait sequences, each one including at least 3 complete gait cycles.

To evaluate the proposed shadow rectification systems using the dataset, this Thesis considers the following protocol. The system is first trained using the sequences belonging to the first session and tests use the sequences of the second session. Next, the system is trained using the sequences belonging to the second session and tested against the sequences of the first session. Finally, the mean of the two results is presented as the recognition accuracy of the proposed systems.



(a) (b)

Figure 7.8: IST shadow gait dataset acquisition setup: (a) camera position, and (b) walking directions.

7.5.2 The IST shadow type dataset

A second new dataset, called the "IST shadow type dataset", has been captured to allow evaluating the performance of the proposed shadow type identification system. This database consists of sequences captured from six individuals across two days. On the first day, individuals are recorded under a clear sky and, thus, the shadows cast by the individuals are sharp – see the right side of Figure 7.9 (b). On the second day, individuals are recorded under a cloudy sky, casting a diffused shadow – see the left side of Figure 7.9 (b). The recordings were conducted at Instituto Superior Técnico, Lisbon, Portugal, on a tennis court, between 10:00 and 12:00, in the month of June 2016, using Xiaomi Redmi 1s cell phone camera. Although this dataset contains video sequences of only 6 individuals, around 300 frames are captured, as each individual walks the full length of the tennis court – see Figure 7.9 (a). Thus, approximately 3600 images are available to evaluate the proposed system. To create the ground truth, all the frames acquired on the first day are indexed as sharp shadows and all the frames acquired on the second day are indexed as diffused shadows. The classification accuracy of the proposed shadow type identification system is obtained by comparing the output of the system to the ground truth.

(a) (b)

Figure 7.9: IST shadow type dataset acquisition setup (a) camera position (b) captured types of shadow.

7.6 Performance evaluation

To evaluate the proposed systems and to compare their performance with the state-of-the-art, the use of the KY IR Shadow dataset [64] is considered in this Thesis. However, since the proposed systems in their current form are not robust to appearance changes, only the first 4 normal gait sequences were used for the evaluation. In addition, unlike the system presented in [64] that uses both shadows and body silhouettes, the proposed systems are evaluated only on the lateral viewpoint shadow silhouettes. The recognition accuracy of the systems is obtained following the 4-fold cross validation protocol presented in [64]. Following the protocol, 3 normal gait sequences are used for training and 1 normal gait sequence is used for testing. The test is repeated until all possible combinations are explored.

Among the systems presented in the state-of-the-art, the use of the feature gait stripe [96], [97], [98], which is computed as the maximum width of the shadow silhouette, does not work well with larger datasets. The use of shadow silhouettes [101] and shadow GEIs [64] operate with a recognition accuracy of 94% and 97%, respectively on the KY IR Shadow dataset, as reported in Table 7.1. The proposed systems perform equally well under such conditions with a recognition accuracy of 94% and 97%, respectively. However, it should be noted that under similar conditions, even the conventional shadow GEI achieves a good recognition accuracy of 92%. The good results can be attributed to the training sequences, which contain the same type of distortions and deformations as observed in the testing sequences. The state-of-the-art systems [101] [64] rely on the same assumption to obtain good recognition results, as illustrated in Figure 7.10 (a, b). Further improvements in the recognition accuracy obtained by such systems can be achieved by employing better performing classifiers.

The proposed systems rectify the shadow silhouettes, compensating for the distortions and deformations present in the shadow silhouettes. However, rectification using TILT can sometimes retain some residual deformations, as illustrated in Figure 7.10 (c), which end up affecting its recognition accuracy, while the proposed 4-point correspondence system always results in rectification of the shadow silhouettes into a canonical view, as illustrated in Figure 7.10 (d). Their performance does not depend on previous knowledge of distortions and deformations in the database, leading to good recognition results even with a simple classifier, as reported in Table 7.1.

Table 7.1: Recognition accuracy (%) of the state-of-the-art and proposed systems that rely on shadow silhouettes.

Recognition systems	Recognition accuracy (%)
Shadow GEI without rectification	92
Shadow GEI AMI system [64]	94
TILT system (7.2)	94
Shadow silhouette AMI system [101]	**97**
4-point correspondence system (7.3)	**97**

As discussed along Chapter 7, under certain conditions the recognition results using body silhouettes can be inferior to the ones obtained using shadow silhouettes. The difference in performance can be attributed to the change in the viewpoint, which affects the body silhouettes, as discussed in Section 2.4.5, while in some conditions the shadow silhouettes can remain relatively unaffected. However, shadow silhouettes are affected by distortions and deformations caused by the camera, which the proposed systems rectify leading to improved recognition results.

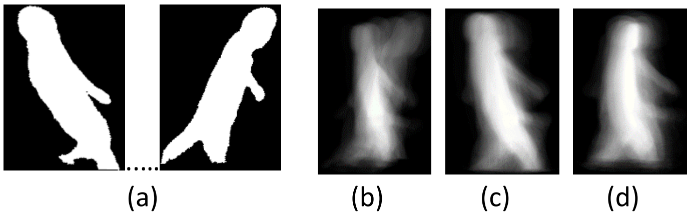


Figure 7.10: Features used by the state-of-the-art and proposed systems (a) shadow silhouettes; (b) GEI, GEI obtained using rectified silhouettes using (c) TILT and (d) 4-point correspondence system.

To highlight the significance of the proposed rectification systems and the advantages of using shadow silhouettes over body silhouettes, the 4-point correspondence system is evaluated using the IST shadow gait dataset, where the three gait cycles obtained along each walking direction (BA and BC) are used to obtain three GEI groups: 1 (start), 2 (middle) and 3 (end). Recognition can then be performed across groups, following the protocol presented in Section 7.5.

Table 7.2: Recognition accuracy (%) of the system using GEIs composed of body silhouettes.

Test set	Training set					
	Group 1 (start)		Group 2 (middle)		Group 3 (end)	
	BA	BC	BA	BC	BA	BC
Group 1	60	62	27	39	05	19
Group 2	31	33	58	68	29	40
Group 3	08	09	28	27	65	73

Table 7.3: Recognition accuracy (%) of the system using GEIs composed of shadow silhouettes.

Test set	Training set					
	Group 1 (start)		Group 2 (middle)		Group 3 (end)	
	BA	BC	BA	BC	BA	BC
Group 1	86	84	43	50	10	11
Group 2	21	13	73	85	43	34
Group 3	10	06	32	30	73	77

In Table 7.2, Table 7.3 and Table 7.4, each entry represents the recognition accuracy along the two considered walking directions: BA and BC – see Figure 7.8 (b). In all cases, diagonal entries in the tables represent the best recognition results, as the testing and the training sequences are obtained from the same part of the gait sequence (start, middle or end). Most state-of-the-art systems rely on such conditions to perform recognition. However, recognition performance when using either body (Table 7.2) or shadow (Table 7.3) silhouettes deteriorates when training and testing is performed using silhouettes from different parts of the gait sequence. The deterioration is caused because the body silhouettes undergo a viewpoint change as the user proceeds along a walking direction, as illustrated in Figure 7.11 (a, b, c). In the case of shadow silhouettes, since the position of the light source (Sun) with respect to the individual along a walking direction remains the same, the shadow silhouettes can be expected to remain almost unchanged, apart from the camera distortions and deformations – see Figure 7.11 (d, e, f). However, such distortions and deformations lead to poor recognition results, as reported in Table 7.3. The problem is addressed by rectifying the shadow silhouettes using the proposed 4-point correspondence system – see Figure 7.11 (g, h, i). The significance of the proposed system is highlighted in Table 7.2 and Table 7.3, where the mean recognition accuracy using the body and shadow silhouette GEIs is 38% and 43%, respectively. Under the same conditions, the proposed 4-point correspondence system performs significantly better, as shown in Table 7.4, as rectifying the shadow silhouettes allows achieving a more consistent performance across all parts of the gait sequence, with a mean recognition accuracy of 75%. Also, given training silhouettes belonging to Group 2 (i.e., the middle part of the gait sequence), only the proposed system performs gait recognition consistently with a mean recognition accuracy of 80%, as reported in Table 7.4.

Table 7.4: Recognition accuracy (%) of the system using GEIs composed of rectified shadow silhouettes obtained using the proposed 4-point correspondence system.

Test set	**Training set**					
	Group 1 (start)		**Group 1 (start)**		**Group 1 (start)**	
	BA	**BC**	**BA**	**BC**	**BA**	**BC**
Group 1	73	80	64	86	45	79
Group 2	71	77	81	89	70	91
Group 3	53	65	74	80	81	90

To evaluate the performance of the proposed shadow type identification system, the Thesis considers the IST shadow type dataset and follows the protocol presented in Section 7.5. The system operates on each frame of the video sequence to identify the type of shadow cast by the individual in it. The system achieves a recognition accuracy of 90%. However, it should be noted that gait features are usually acquired from an integer number of complete gait cycles, each with a typical duration between 1 and 2 seconds. Thus, using a voting policy the shadow type can be identified for the entire duration of the gait cycle, further improving the recognition accuracy of the proposed system to almost 100%.

The proposed system cannot tackle situations where the individuals' attire blends in with the background creating a camouflage effect. Due to the lack of contrast under such conditions, the individuals' body region is represented with high intensity ratio values in the histogram plot. Thus, the proposed system fails to identify the type of shadow under such conditions.

In the results obtained, it is observed that some sharp shadows are falsely identified as diffused shadows, as the histograms representing the intensity ratio values of the individual and the shadow detect two peaks. However, under such conditions, the distance between the two peaks is extremely small, when compared to the diffused shadows. Thus, by applying a threshold value to the distance between the two peaks, the system's accuracy can be easily improved.

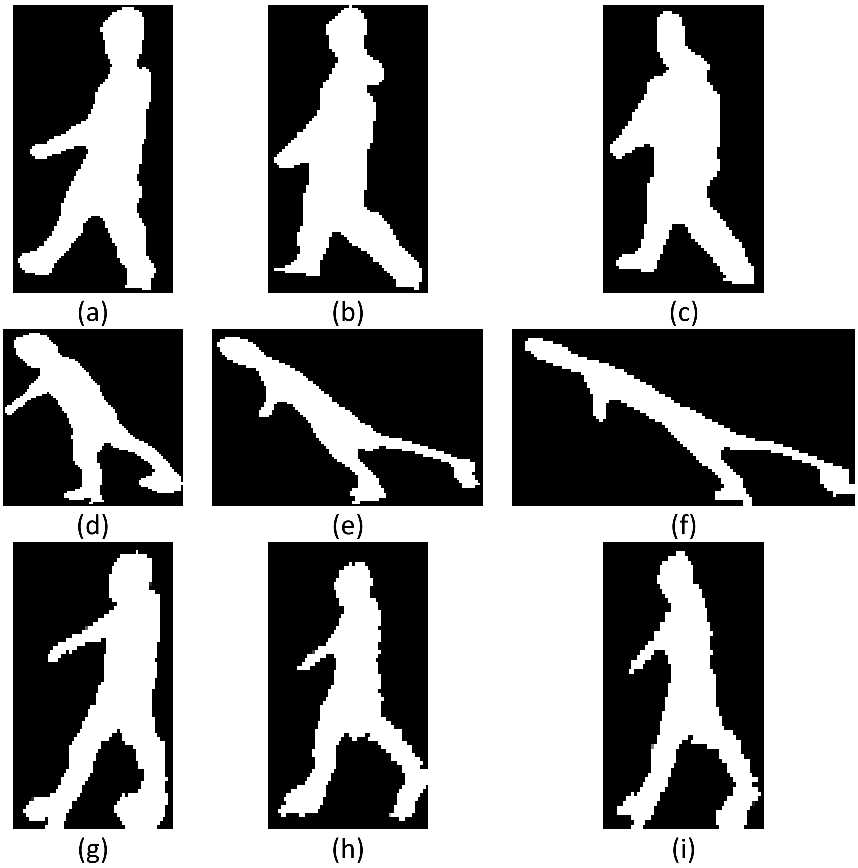


Figure 7.11: Silhouettes at the start, middle and end of a gait sequence. (a, b, c) represent body silhouettes; (d, e, f) represent shadow silhouettes and (g, h, i) rectified shadow silhouettes.

Part III: Gait-based pathology classification

8 PROPOSAL FOR ACCURATE ESTIMATION OF TEMPORAL BIOMECHANICAL GAIT FEATURES

8.1 INTRODUCTION

Gait analysis for medical diagnosis, as discussed in Section 3.3, is typically performed using systems such as the motion capture system [128], or a force plate-based system [129], for their reliability and accuracy. However, these systems are expensive to install and are inaccessible to most individuals in a daily life setting as they can operate effectively only in dedicated laboratories, under the supervision of trained professionals. Consequently, the use of such systems for the assessment of the individuals' gait on a regular basis may not be easy. Alternatively, using systems based on IMUs can address some of the identified shortcomings [130]. However, since IMU devices must be mounted on distinct and specific locations of the individuals' body, trained professionals are required to obtain the best results.

To obviate the limitations, 2D vision-based systems can be used to perform gait analysis in a non-invasive manner, as discussed in Section 3.3.2. Nevertheless, most of the 2D vision-based systems reported in the literature focus on the classification of gait as either normal or impaired, and only a limited number explore the temporal gait features for a more detailed analysis, notably to help the medical diagnosis. Additionally, among the latter group of systems, the accuracy of the features obtained is seldom validated for clinical use. This Thesis presents a novel system to obtain temporal gait features, including stance time, swing time and gait cycle time, by detecting the initial contact (IC) and toe off (TO) events during a gait cycle. The system's usage for clinical assessment is also validated, by comparing the obtained results against those obtained when using a motion capture system, which is considered as the gold standard in clinical assessment [128].

8.2 ESTIMATION OF STANCE TIME, SWING TIME AND GAIT CYCLE TIME USING IC AND TO

A novel system is proposed in this Thesis that estimates a set of biomechanical gait features that are frequently used in the clinical evaluation of individuals suffering from some type of gait related pathology. Temporal gait features are estimated by first detecting the occurrence of the IC of a foot with the floor and the instant of TO, which occurs when the foot leaves the floor, during a gait cycle.

The IC and TO are detected by first identifying the position when the foot is in complete contact with the floor, as illustrated in Figure 8.1. For this purpose, the proposed system creates a "foot flat image" using a GTI. As discussed in Section 5.3, the GTI highlights the feet positions of the individuals due to the overlap between their silhouettes – see Figure 8.2 (a). The overlap among the silhouettes is highest at locations that remain static over a long period, i.e., the foot flat positions during the gait cycle, which can thus be identified in the proposed system by applying a threshold to the GTI. To prevent the selection of other body parts, only the bottom 10% of the individual silhouettes in the GTI is considered – see Figure 8.2 (b). The threshold value for each foot flat location can be selected using the highest intensity values along the walking direction, which can be identified following the steps presented in Section 5.3. As illustrated in Figure 8.2 (c), the proposed system uses the valleys in the intensity plot to separate two neighbouring foot flat positions, followed by the application of Otsu's thresholding [131], to obtain the corresponding foot flat positions. The resulting image, called the “foot flat image”, contains the foot flat locations for a given gait sequence – see Figure 8.2 (d).

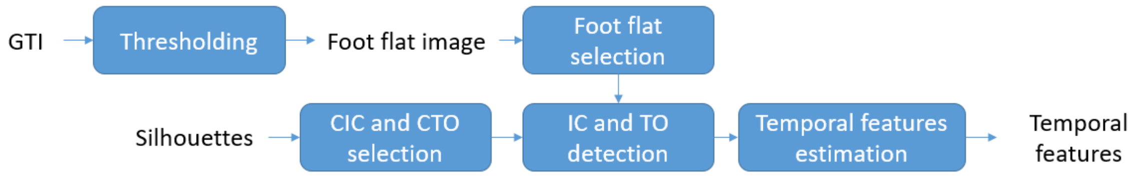


Figure 8.1: Architecture of the proposed temporal features estimation system.

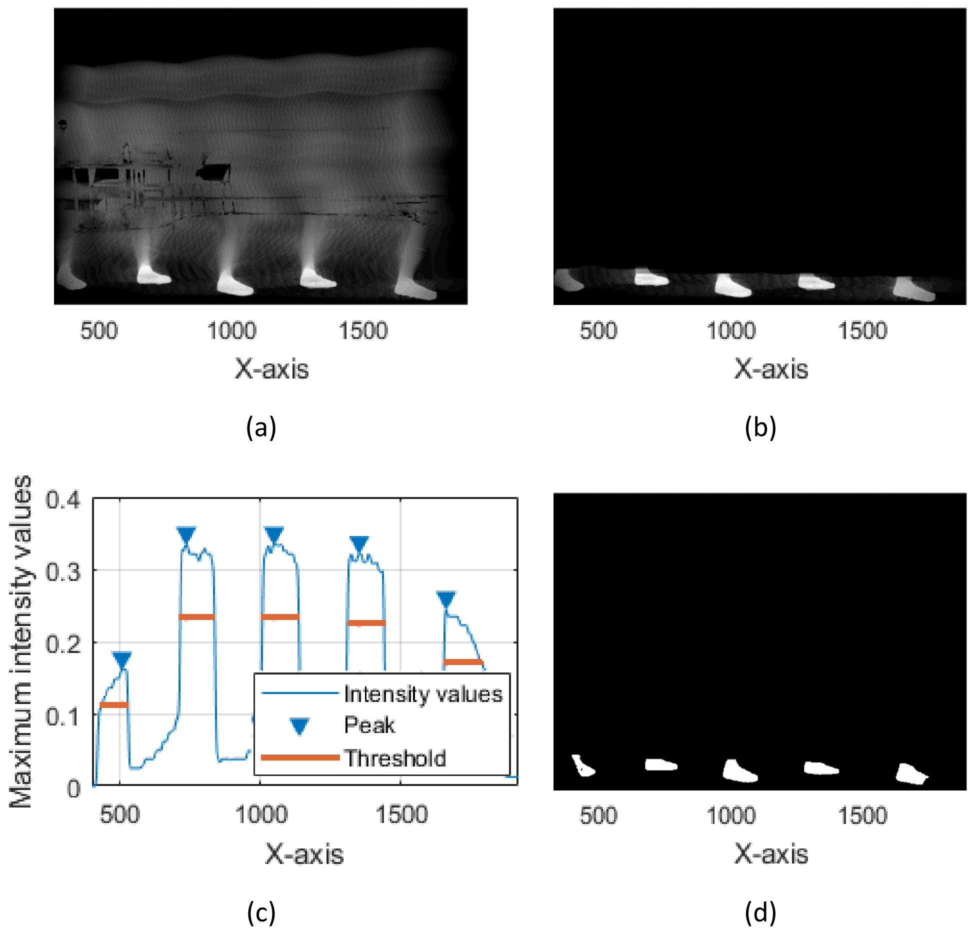


Figure 8.2: Illustration of the proposed foot flat identification steps: (a) GTI, (b) selected feet region of the GTI, (c) thresholding, (d) resulting “foot flat image”.

After foot flat detection, the proposed system can analyse the overlap between the foot flat and the silhouettes obtained from a video sequence to estimate the IC and TO events. However, to avoid false positives, the proposed system first identifies two sets of silhouettes in which each of the two events are most likely to occur. The two sets of silhouettes are selected near every double support phase of the gait cycle, which corresponds to the moment when both feet are in contact with the floor, as candidates for initial contact (CIC) and candidates for toe off (CTO) respectively. The CIC set includes silhouettes that lie between the silhouette with the mean width value and the next silhouette with the maximum width, as the IC instant that marks the start of double support phase occurs somewhere in-between. Similarly, the CTO set includes silhouettes that lie between the silhouettes with maximum width value and the next silhouette with the mean width, as illustrated in Figure 8.3. The width of the silhouette is normalised by subtracting each width value by the mean and dividing it by the standard deviation.

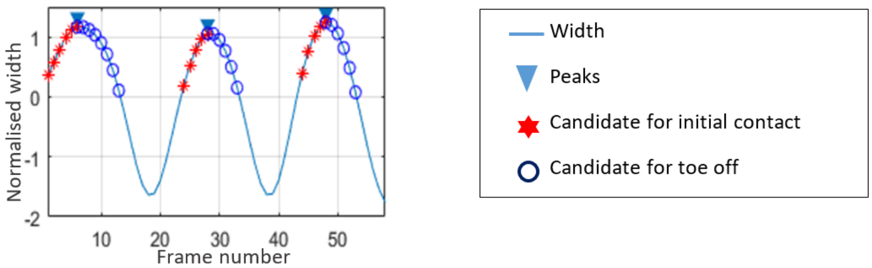


Figure 8.3: Selection of candidate frames for the estimation of initial contact and toe off.

The IC occurs at the individuals' leading foot, while the TO occurs at the individuals' trailing foot. Thus, the proposed system selects only the leading foot from every CIC set and the trailing foot from every CTO set, as illustrated in Figure 8.4 (b).

As illustrated in Figure 8.4, for a given gait sequence, two double support phases will be detected per each observed gait cycle. For every double support phase i, a set of CICs, CIC_i^n, and CTOs, CTO_i^m, can be obtained, where n, m are the index of candidates in each set. The estimation of the occurrence of the IC and the TO instants can then be performed by analysing the overlap between a foot flat and the selected foot from the set of CICs and CTOs, respectively – see Figure 8.4 (d).

As illustrated in Figure 8.4 (b, c) for k double support phases, there exist k foot flats in the foot flat image. Thus, the selection of the foot flat and the corresponding set of CICs and CTOs, to estimate the IC and the TO can be done as follows.

For CIC_i^n and CTO_i^m corresponding to double support phase i, the foot flat i is selected to estimate the TO, and foot flat $i + 1$ is selected to estimate IC. As an example, considering Figure 8.4, for CIC_2^n and CTO_2^m corresponding to the second double support phase, foot flat 2 is selected to estimate TO and foot flat 3 is selected for estimating IC. The overlap between the foot flat 3 and CIC_2^1, CIC_2^6 and CIC_2^{10} and the flat foot 2 and CTO_2^1, CTO_2^5 and CTO_2^9 are illustrated in Figure 8.4 (d), highlighting the instants where the IC and TO occur with red boxes. Also, in Figure 8.4 (d), white represents overlapping areas, while grey represents the remaining parts of the feet.

For the selected set of CIC/CTO and the corresponding foot flat, the overlap can be measured as the percentage of CIC/CTO foot covered by the selected foot flat. As illustrated in Figure 8.4 (e), for the IC, the percentage of overlapped area increases rapidly at the start and tends to stabilize at the end. Thus, the initial contact can be estimated as the point where slope of the increase in the overlapped area slows down significantly. It is observed that such slow down occurs when almost 80% of the CIC foot is overlapped. Thus, the proposed system estimates the IC as the first frame for which the overlap of the CIC foot and the foot flat exceeds 80%. Similarly, the toe off can be estimated by analysing the overlap between the feet from the set of CTOs and the selected foot flat. The toe off can be estimated as the first frame for which the overlap between the CTO foot and the selected foot flat becomes zero – see example in Figure 8.4 (f).

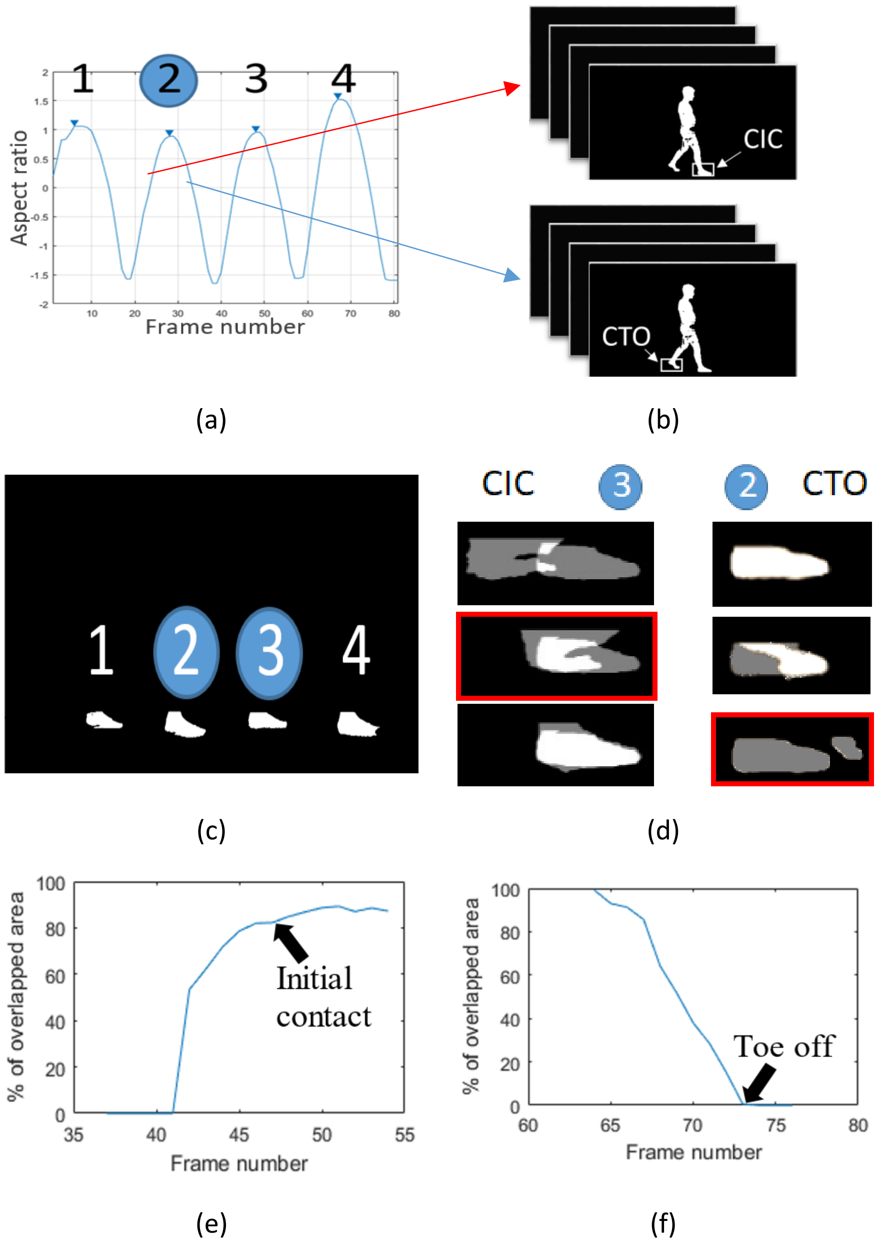


Figure 8.4: Examples of intermediate steps in IC and TO estimation: (a), (b) selection of foot from CIC and CTO, (c) flat feet selection, (d) overlap between flat foot and CIC/CTO feet, estimation of: (e) Initial contact (f) toe off.

The result from the previous step is a sequence of frame numbers indicating the estimated instances of IC and TO. To have a better understanding of the individuals' gait, frame numbers can be converted into timestamps using the frame rate of the video sequence. The timestamps can then be used to estimate additional temporal gait features, such as the left and right stance times, left and right swing times and the left and right gait cycle times.

8.3 IST-KUL DATASET ACQUISITION

The proposed system estimates the temporal gait features using silhouettes obtained from a single 2D video sequence. To evaluate its performance, and validate the proposed system, this Thesis presents a new dataset, called "IST-KUL gait dataset", which consists of ten individuals (7 male and 3 females between the ages of 20-25) each recorded five times, using a Casio Exilim

EX-ZR100 camera, which has a spatial resolution of 1920×1080 and captures video with a frame rate of 30 fps. The individuals are recorded walking in a perpendicular direction with respect to the camera axis (i.e., a side viewpoint is recorded), as illustrated in Figure 8.5. Everyone is also simultaneously recorded using a motion capture system, installed in a laboratory of the Rehabilitation Sciences Group of the Bruges Campus of KU Leuven, consisting of six infrared Optitrack Flex 13 cameras, each with a resolution of 1280×1024 and a maximal frame rate of 120 fps. All cameras are synchronized and calibrated using Motive Tracker v.01.90. The individuals' gait is captured using forty-four reflective markers secured at carefully selected locations on their bodies, according to the lower limb and trunk model [128], to estimate the gait features. The gait features are obtained using the marker and coordinate based algorithm presented in [128]. The motion capture system is considered as the gold standard for the estimation of biomechanical gait features [128].

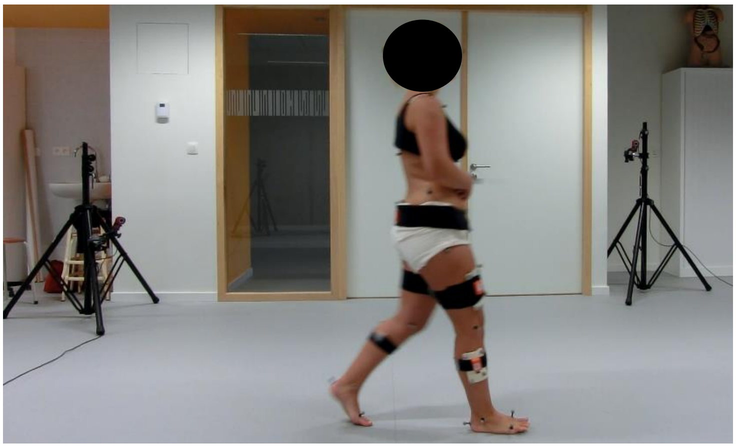


Figure 8.5: Camera setup for the IST-KUL dataset.

8.4 PERFORMANCE EVALUATION

The proposed system detects the IC and the TO events to estimate various temporal gait features. Before evaluating the accuracy of the features obtained, the IC and the TO detection accuracy of the proposed system is evaluated using the CASIA A [106] dataset. Following the protocol presented in [54], the system is evaluated using the 0°, lateral viewpoint sequences, of 20 individuals captured on 4 different days. Before the evaluation, the frames corresponding to the IC and TO events are manually annotated to construct the ground truth. Considering the camera frame rate of 30 fps, and since normal gait cycles usually last for 1-2 seconds, it is observed that in several cases the exact instant of IC or TO is not recorded. Thus, the ground truth is constructed considering the frame just before or the frame just after the event. The decision is made by visually inspecting each frame. Therefore, for evaluation purposes, the system can be allowed an error margin of ±1 frame with respect to the ground truth.

The proposed system is evaluated by analysing the distance between its output and the ground truth. As illustrated in Figure 8.6, it estimates IC and TO with an accuracy of 72% and 88%, respectively. The TO estimation is significantly better than the IC, as it is very easy to detect zero overlap between the foot flat and the foot from the CTO set using the proposed system.

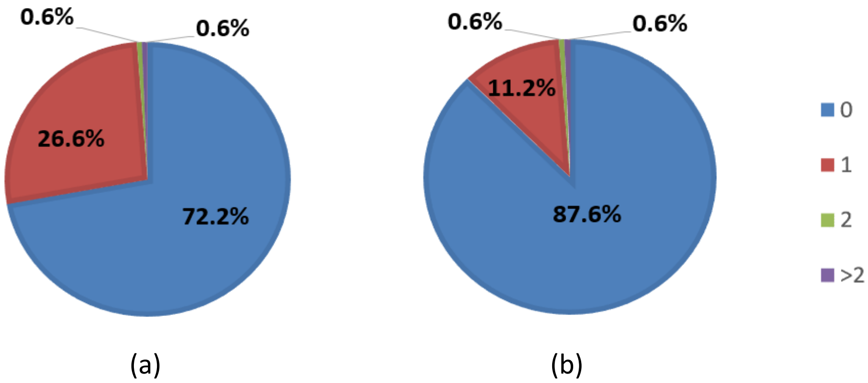


Figure 8.6: Percentage of correctly estimated (a) ICs and (b) TOs w.r.t. the distance from the ground truth.

When allowing an error margin of ±1 frame w.r.t. the ground truth, the detection accuracy of IC and TO improves significantly, to 99%. Using an error margin of ±2 frames, which is adopted by most state-of-the-art systems [54], [56], the accuracy further improves to almost 100%, which is significantly better than achieved by the state-of-the-art systems, as reported in Table 8.1. A root means square error (RMSE) is also computed using the distance between the frame numbers estimated by the proposed system and the ground truth, which confirms that the proposed system performs significantly better than the state-of-the-art. It should also be noted that the systems presented in [54], and [56] ignore the IC and TO occurring at the beginning and the end of the gait sequences due to the presence of incompletely segmented silhouettes. The results presented in Table 8.1 report these ignored events as a part of the failed detections. The proposed system, unlike the state-of-the-art, can operate even on such silhouettes.

Table 8.1: IC and TO estimation with an error margin of ±2 frames w.r.t. the ground truth.

	System [56]		System [54]		Proposed System (8.2)	
	IC	**TO**	**IC**	**TO**	**IC**	**TO**
Correct Estimations	316	315	319	305	**332**	**330**
False Estimations	8	5	9	8	**0**	**1**
Failed Detections	10	12	6	19	**2**	**1**
RMSE	0.98	0.95	0.89	1.62	**0.54**	**0.42**

To evaluate the performance of the proposed system when operating on gait sequences acquired from different viewpoints, 10 individuals from CASIA B dataset are considered. The evaluation is performed using only a subset of the dataset, as the ground truth needs to be created by visually inspecting and manually annotating each video frame. The CASIA B dataset contains 6 normal sequences of the individuals, captured along 11 different viewpoints – see Section 3.2.4. However, for the current evaluations only the 7 viewpoints between 36˚ to 144˚ are considered, because for the rest, the overlapping silhouettes lead to poor foot flat detection. Considering an error margin of ±1 frame w.r.t. the ground truth, the proposed system detects the IC and TO with an accuracy of almost 100%, as reported in Table 8.2. This is a significant improvement over the state-of-the-art systems [54], [56] that operate only on lateral (90˚ viewpoint) sequences.

Table 8.2: IC and TO estimation with an error margin of ±1 frame across different viewpoints.

View (˚)	Correct Estimations (%)		False Estimations (%)		Failed Detections (%)	
	IC	TO	IC	TO	IC	TO
36	100	99	0	1	0	0
54	100	99	0	1	0	0
73	100	100	0	0	0	0
90	100	100	0	0	0	0
108	100	100	0	0	0	0
126	100	100	0	0	0	0
144	100	100	0	0	0	0

A second set of evaluations is performed to analyse the possibility of using the proposed system for clinical assessment. These evaluations are conducted using the IST-KUL gait dataset, to compare the results against the motion capture system [128], which is considered as the gold standard in clinical assessment. The evaluation involves estimating the temporal features from the 2D sequences available in the IST-KUL gait dataset using the proposed system. It is observed that using the error margin of ±1 frame, the proposed system correctly estimates the IC and the TO with 99% accuracy, when compared to the manually annotated ground truth. It should be noted that the proposed system attempts to perform the same type of assessment as the motion capture system [128], but without any calibrations or initial setup and with a single camera. The camera also operates at approximately one forth the framerate of the motion capture system [128].

Table 8.3 reports the obtained results, where "L" and "R" represent the left and right legs, respectively. The agreement between the two systems is calculated using an intra class correlation coefficient (ICC). Notice that according to Fleiss and Cohen [132], an ICC in the ranges 0.00 – 0.39 is considered 'bad', 0.40 – 0.73 'moderate', 0.74 – 0.90 'good' and 0.91 – 1.00 'excellent'.

As reported in Table 8.3, the 'good' to 'excellent' ICC values obtained suggest a high level of agreement between the results of the proposed system and the optoelectronic system [128]. For the left and right gait cycle times, a small proportional bias (the difference between the means obtained by the two systems) of 0.02 sec is observed. The ICC indicates that the proposed system can estimate IC with a high level of accuracy. A slightly higher bias is found when evaluating the left and right swing and stance times. However, a good correlation can still be observed between the two systems. The evaluation also provides reliability value for the observed correlation, called the p-value. A score of less than 0.001 for the p-value of every entry in Table 8.3 also suggests a very strong evidence for the observed correlation.

From the obtained results, it can be concluded that the proposed system is a viable alternative to the motion capture system [128]. More importantly, the proposed system can operate in unconstrained environments, thus allowing a more frequent follow up of the individuals in between visiting the laboratory, where the motion capture system is installed.

Table 8.3: Comparison between the proposed and motion capture system.

	Proposed System (8.2)		Optoelectronic System [128]		ICC	Mean Diff	SD (sec)	RMSE (sec)
	Mean (sec)	**SD (sec)**	**Mean (sec)**	**SD (sec)**				
Stance Time L	0.77	0.08	0.72	0.07	0.85	0.05	0.06	0.06
Stance Time R	0.80	0.09	0.72	0.08	0.83	0.08	0.06	0.09
Swing Time L	0.42	0.05	0.47	0.05	0.81	-0.05	0.04	0.05
Swing Time R	0.45	0.05	0.47	0.06	0.83	-0.03	0.04	0.04
Cycle Time L	1.22	0.13	1.20	0.11	0.92	0.02	0.07	0.05
Cycle Time R	1.21	0.12	1.20	0.12	0.92	0.02	0.06	0.05

9 PROPOSAL FOR CLASSIFICATION OF GAIT RELATED PATHOLOGIES

9.1 INTRODUCTION

The second problem addressed in terms of medical diagnosis is related to the classification of impaired gait, and to the possibility of identifying the corresponding gait pathology. Most state-of-the-art 2D vision-based systems perform only a binary classification of gait as being either normal or impaired, as discussed in Section 3.3. A further classification of gait pathologies can provide a preliminary assessment of the type or the severity of the observed gait disorder. Supporting the latter possibility using a simple system based on a single 2D camera would make such assessment accessible to individuals in daily life settings, where the constant presence of trained professionals is not possible.

Some of the existing 2D vision-based systems can distinguish between different types of gait pathologies using gait representations such as the GEI [28]. However, they perform poorly when trying to discriminate certain types of pathologies, such as diplegia and Parkinson's diseases. Thus, this Thesis presents two novel systems able to perform classification across different gait related pathologies. The first system explores "hand-crafted" biomechanical features, such as the well-known step length, speed, torso orientation and shift in centre of gravity (COG), and proposes additional novel features, such as the foot flat ratio, normalized step count and the amount of movement while walking, to classify the different gait related pathologies. The second system uses a fine-tuned VGG-19 deep convolutional neural network (CNN) [133] to obtain features from the GEIs that best represent the different gait pathologies. Both proposed systems can perform well even under self-occlusions (i.e., when the part of the body closer to the camera occludes other parts of the body), which can be difficult to handle in certain pathologies for which one foot of the individual remains occluded by the other throughout the gait cycle.

9.2 CLASSIFICATION OF PATHOLOGIES USING "HAND-CRAFTED" BIOMECHANICAL FEATURES

The first proposed system uses the silhouettes obtained from the lateral viewpoint of the walking individuals to estimate "hand-crafted" biomechanical gait features, which can then be used to classify gait sequences across different gait related pathologies. The proposed system architecture is illustrated in Figure 9.1. The features used can be classified into two groups:

- **Feet related features**: These features are obtained using the spatio-temporal information acquired from the individuals' feet area. They include well-known left and right step lengths, step length symmetry, speed, and novel left and right foot flat ratios and the normalized step count.
- **Body related features:** These features are obtained using the information acquired from the entire body of the individual. They include the novel amount of movement in the left and the right side of the body, movement symmetry and the shift in the COG with respect to its centre of support (COS), and torso orientation.

The proposed system estimates the step length as the distance covered between the initial contacts of the observed foot and the other foot. In healthy individuals, detecting the initial contact, and thus estimating the step length can be easy as the feet are separated by some distance while walking, as illustrated in Figure 9.2 (a) [54]. However, due to certain pathologies, the stride of the individuals can be extremely short, leading to self-occlusions as illustrated in Figure 9.2 (b). Under such conditions, identifying the exact instant of initial contact or estimating the step length using silhouettes can be difficult. The proposed system tackles this problem by detecting the "foot flat" positions during the gait cycle. A gait cycle includes two foot flats, one for each foot, occurring right after the initial contact. To minimize the effect of self-occlusions, foot flat instances are detected by analysing half gait cycle at a time, i.e., the span between two consecutive initial contact events (of opposite feet). However, since determining the exact instant of initial contact is difficult, the proposed system approximates it as the instant in time for which the distance between the two feet is maximum [118], as illustrated in Figure 9.3.

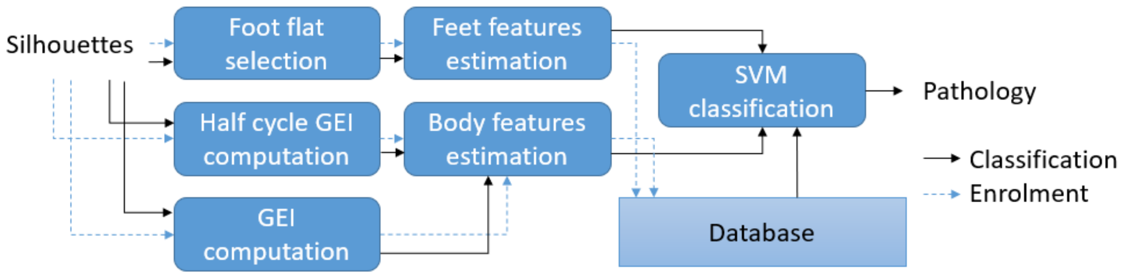


Figure 9.1: Architecture of the proposed gait-based pathology classification system, using "hand-crafted" biomechanical features.

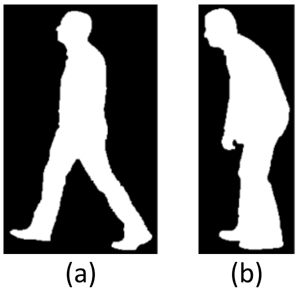


Figure 9.2: Silhouettes belonging to a healthy individual and (b) an individual suffering from Parkinson's diseases.

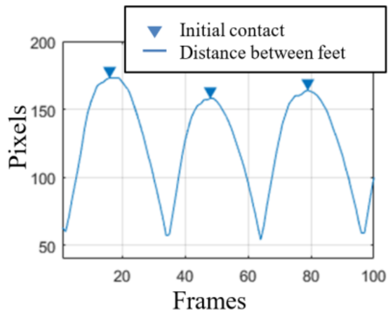


Figure 9.3: Plot representing the distance between feet for a gait cycle.

To obtain foot flat positions, the proposed system uses the feet region of the silhouettes to obtain an average feet image (AFI). The feet region can be selected as the lower 10% fraction of the silhouettes [134]. The $AFI(x, y)$ can then be computed by averaging the resulting T feet silhouettes images, $I_{feet}(x, y, t)$, available between two initial contacts, according to (9.1), and illustrated in Figure 9.4 (a) and (b).

$$AFI(x, y) = \frac{1}{T}\sum_{t=1}^{T} I_{feet}(x, y, t) \quad (9.1)$$

The AFI highlights the foot when it is in complete contact with the ground. Thus, by applying the Otsu thresholding [131] to the AFI, the proposed system can estimate the position of the foot flat, as illustrated in Figure 9.4 (c), which can then be used to compute the various feet related features.

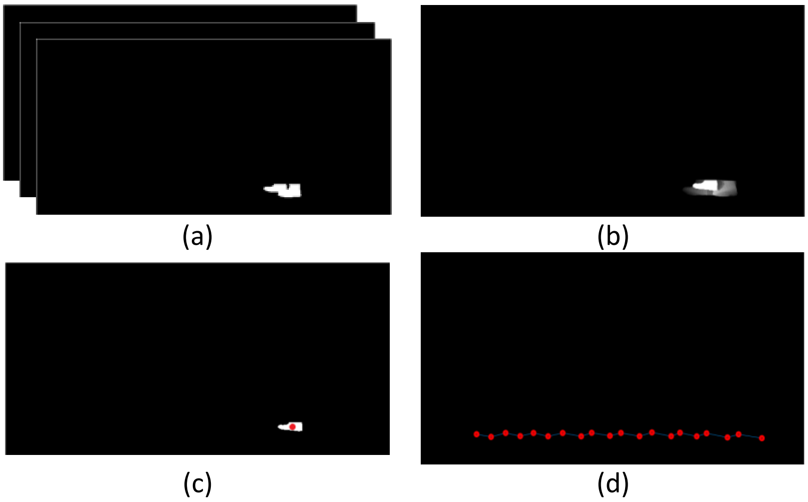


Figure 9.4: Intermediate steps for step length estimation: (a) segmented feet silhouettes between two initial contacts; (b) AFI obtained by averaging the feet silhouettes, (c) position of the foot flat obtained by applying a threshold, (d) centroids of foot flats obtained for the entire video sequence.

The proposed system measures the step length (SL) using the foot flat positions obtained from the entire sequence. However, due to the lack of depth information, there can be a significant difference between the scales of the two feet. Thus, to minimize errors, the proposed system computes the centroid of each foot flat and measures the Euclidean distance between two consecutive centroids as the step length. Knowing the walking direction of the individuals with respect to the camera, the proposed system can identify the foot closer to the camera as the right or the left foot. Thus, the proposed system can estimate both the left and right step lengths. To identify which foot is closer to the camera, since depth information is not available, the foot flat centroid positions can be used. As illustrated in Figure 9.4 (d), the centroid of the foot further away from the camera appears at a more elevated position in the image. Thus, by comparing the y-coordinate of the centroids, the step lengths can be classified as either left or right step lengths. Since the video sequence contains multiple gait cycles and thus allows computing multiple feature values, a median is computed to increase the proposed system's robustness to outliers [134]. Therefore, the proposed system computes the median of the left and right step lengths as SL_i^{left} and SL_j^{right}, where i and j represent the indices of the left and right step lengths respectively. The distance is measured in pixels but, since the silhouettes are normalized with respect to height, while maintaining their original aspect ratio, the system is robust to scale changes. A step length symmetry score, SL_{symm}, can then be computed as the absolute difference between the medians of left and the right step lengths, according to (9.2). The symmetry score is consistent across different individuals, observed at different distances from the camera due to the normalisation step.

$$SL_{symm} = \left| median\left(SL_i^{left}\right) - median\left(SL_j^{right}\right) \right| \tag{9.2}$$

The proposed system can also compute the normalized step count and speed of the individuals' movement using the foot flat information. Normalized step count (NSC) is computed as the total number of foot flat instances k, divided by the total distance travelled, according to (9.3). The distance travelled is measured as the length summation of the n observed step lengths.

$$NSC = \frac{k}{\sum_{i=1}^{n} SL_i} \tag{9.3}$$

The speed (S) of an individual's movement is computed by dividing the total distance travelled by the duration of the video sequence, according to (9.4). The duration of the video sequence, d (in sec), is measured between the first and last IC.

$$S = \frac{1}{d}\sum_{i=1}^{n} SL_i \tag{9.4}$$

A video sequence of the walking individual is composed of several repetitions of a gait cycle, delimited by the IC of the observed foot. It is also possible to divide each gait cycle into two phases separated by a TO event. The phase before the TO is called the stance phase, while the phase following the TO is called the swing phase. Using the IC and the TO, the proposed system can estimate the duration of the stance phase. As discussed in [118], the duration of the stance and swing phases are not unique enough to distinguish between different types of gait pathologies. However, the duration of foot flat during the stance phase, can change significantly depending on the type of gait pathologies. Thus, the proposed system computes a "foot flat ratio" (FFR) feature, which can be defined as the fraction of the stance phase during which the foot flat occurs.

To compute the FFR, the proposed system measures the amount of overlap between the foot flat and the silhouettes belonging to the corresponding stance phase. It estimates the foot flat duration by counting the number of frames for which the foot flat is completely covered by the silhouettes – see Figure 9.5. Foot flat ratio values, for both the left and right feet, can then be computed according to (9.5)

$$FFR = \frac{flat\ foot\ duration}{stance\ phase\ duration} \tag{9.5}$$

Apart from the feet, several biomechanical features can also be estimated from the body silhouettes, which can characterise some gait pathologies that an individual may be suffering from. For example, an individual's movements can get severely restricted and the posture of the individual can be severely altered due to disorders such as Parkinson's disease [28]. Thus, a measurement of the amount of movement and posture instability can be useful for classifying such gait pathologies. In addition, in some cases, the movement of a single limb may be restricted, or more restricted than the other limb. To collect information about the movement of each leg, analysing the movement for every half gait cycle can be useful.

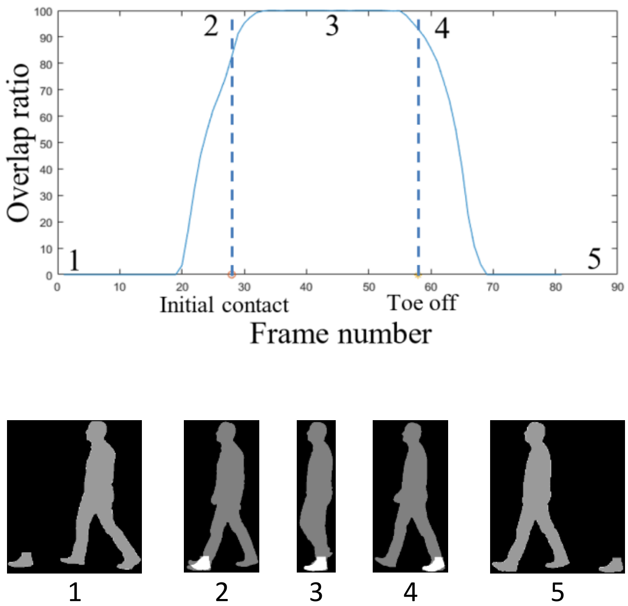


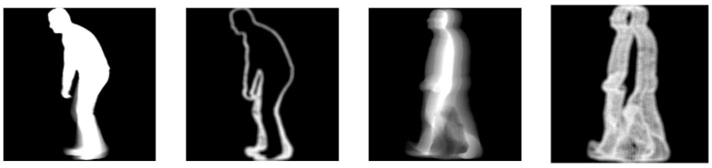
Figure 9.5: Plot representing the foot flat overlap ratio (top) and the corresponding foot and walking individual's silhouettes (bottom).

The proposed system computes the entropy during every half gait cycle as a measure of the amount of movement. However, unlike what is done for feet related features computation, here the half gait cycle is delimited by the mid-stance and mid-swing events, as it contains the part of the gait cycle where individuals shift the body weight from one side of the body onto the other. The proposed system can thus capture movement restrictions while shifting weight onto the impaired side of the body. The mid-stance and mid-swing instants are approximated as the instants of the gait cycle when the distance between the two feet is minimum, corresponding to the valleys represented in Figure 9.3. The silhouettes belonging to each half gait cycle, numbered from 1 to N, can then be cropped, $I_c(x, y, n)$ and averaged to obtain the half cycle GEI, $GEI_{hc}(x, y)$, according to (9.6).

$$GEI_{hc}(x, y) = \frac{1}{N} \sum_{n=1}^{N} I_c(x, y, n) \tag{9.6}$$

The amount of movement (AOM), can then be computed over the half cycle GEI according to (9.7), where P_i is the probability calculated using the grey level histogram of the half cycle GEI for each pixel value i. As illustrated in Figure 9.6 (b, d), the restriction in movement can be effectively represented using Shannon entropy [135].

$$AOM = -\sum_{i} P_i log_2 P_i \tag{9.7}$$



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

Figure 9.6: Half cycle GEI computed using (a) impaired and (c) healthy gait silhouettes, and (b, d) the corresponding GEnI representations, where entropy highlights the amount of movement.

Following the foot flat, the amount of movement features can also be classified into left, AOM_i^{left} and right, AOM_j^{right}, where i and j represent the indices of the left and right half gait cycles. The classification is performed according to the foot that enters an initial contact during the considered half gait cycle. A symmetry measure, AOM_{symm} can then be computed to represent the difference in movement between left and right sides, according to (9.8).

$$AOM_{symm} = \left| median\left(AOM_i^{left}\right) - median\left(AOM_j^{right}\right) \right| \tag{9.8}$$

As illustrated in Figure 9.7, certain types of gait pathologies can affect the posture of individuals, being reflected as a change in the orientation of the torso and therefore as a shift in the individuals' COG with respect to the COS. Healthy individuals walk such that their COG and COS are always approximately vertically aligned. The proposed system computes the amount of shift using a GEI computed similarly to (9.6), but over the entire gait cycle. Using the GEI provides robustness to variations in the shift in COG occurring at different instants of the gait cycle. The COG is measured as a weighted centroid of the GEI, using the GEI intensity values as weights. The COS is measured as the centre of the feet region of the GEI, obtained by segmenting the bottom 10% of the GEI as feet using a human anatomy ratio [134]. The shift in COG, COG_{shift} can then be computed as the absolute difference between the horizontal coordinates of the COG and COS, according to (9.9)

$$COG_{shift} = |COG_x - COS_x| \tag{9.9}$$

The last feature considered, called orientation of the torso (OT), is also computed using the GEI obtained from a complete gait cycle. The proposed system selects the torso as the top 50% of the GEI, according to the human anatomy ratios presented in [134]. It then performs principal component analysis over the torso and measures the angle between the horizontal axis and the first principal component, $PC(PC_x, PC_y)$, according to (9.10).

$$OT = \left| tan^{-1}\left(\frac{PC_y}{PC_x}\right) \times \frac{180}{\pi} \right| \quad (9.10)$$

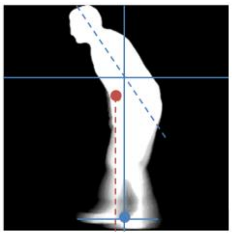


Figure 9.7: GEI highlighting shift in COG (middle silhouette point in red) with respect to the COS (lower silhouette point in blue) and the orientation of the torso.

The proposed system can perform classification in two different ways. First, the system can use each feature to classify gait as either normal or impaired. Although an individual feature can be used to classify the gait, the type of pathology that can be detected may vary depending on the features used. For instance, features such as the shift in COG or the OT can help detect posture instabilities, features such as SL or AOM can be used to detect asymmetric gait, while features such as FFR, NSC or S can be used to detect slow movements and other deviations. Thus, using all the available features together can allow the proposed system, not only to detect impaired gait, but to further classify gait based on the type of pathology. It might even allow determining the severity of the observed disorders. For example, Parkinson's disease reduces the walking speed, alters the posture and restricts the movement of an individual, while disorders such as hemiplegia restrict the movement of a single side of the body. The proposed system can distinguish between gait pathologies using the proposed biomechanical features. It is also possible to identify the side of the body whose movement is affected, such as in the case of hemiplegia, as the considered features allow differentiating between left and right side impairments. The proposed system performs such classifications using an SVM, a discriminative classifier that separates data using a hyperplane [136]. To improve the classification accuracy of the proposed system, the SVM is used with a quadratic kernel.

9.3 Classification of pathologies using a deep CNN and transfer learning

The performance of most 2D vision-based systems that classify gait across different gait related pathologies, depend on the quality of silhouettes used - see Section 3.3. Thus, poor segmentation of the silhouettes can significantly reduce their classification accuracy. Since gait representations used for biometric recognition, such as the GEI, are robust to such limitations [28], they can provide a more reliable analysis of the individuals gait, allowing operation even in less constrained settings where silhouettes are expected to contain segmentation errors, for instance due to video acquisition against a dynamic background. Apart from gait representation, the performance of the pathology classification systems can also be improved by using better performing classifiers. Recently, the use of deep CNNs, such as VGG-16 [137], pose-based temporal-spatial networks [138], have shown a significant improvement in the performance of silhouette-based gait recognition systems. Similar improvements have also been seen in the medical domain, especially in detecting Alzheimer's disease [139]. Thus, this Thesis explores the use of deep learning techniques along with GEI, to improve the performance of the gait-based pathology classification systems.

The proposed system uses GEIs as input to a fine-tuned VGG-19 [133], to obtain features to perform classification of gait across different gait related pathologies. VGG-19 is selected as it is the best performing deep CNN for pathology classification – see Section 9.4. VGG-19 is a 19-layer CNN [133] that can be considered as a stack of convolutional layers, as illustrated in Figure 9.8, with a filter size of 3×3, with stride and pad of 1, along with max pooling layers of size 2×2 with stride of 2. The convolutional layers in VGG-19 detect the local features available in the input GEI. Next, the max pooling layers reduce the size of the feature vectors obtained, thus reducing the computational complexity of the network. A series of such layers is followed by two fully connected layers that learn the non-linear relationship among the local features. The final layer of the network is a softmax layer, which performs classification.

The starting point for development has been the VGG-19 trained on ImageNet [133], which can classify images across 1000 different image groups. That model has been trained using over 1.3 million images. However, a dataset of such a scale containing sequences of gait affected by different gait pathologies is currently unavailable. In addition, training VGG-19 with a small dataset is expected to result in problems, such as overfitting to the small training set. This limitation can be addressed using transfer learning, a machine learning technique where a model trained to address one problem is re-purposed to address a second related problem. The proposed system therefore uses the VGG-19 model pre-trained on ImageNet [133] and part of the network is retrained with gait GEIs from an available dataset, to fine-tune the model parameters. In a deep CNN, the initial convolutional layers of the network typically detect simple features, and features become more complex in subsequent network layers, with final layers capturing more problem specific features. Thus, when using transfer learning, the final layers of the network can be re-trained such that they are fine-tuned for the classification of gait pathologies. The details of the fine-tuning process are discussed in Section 9.4.

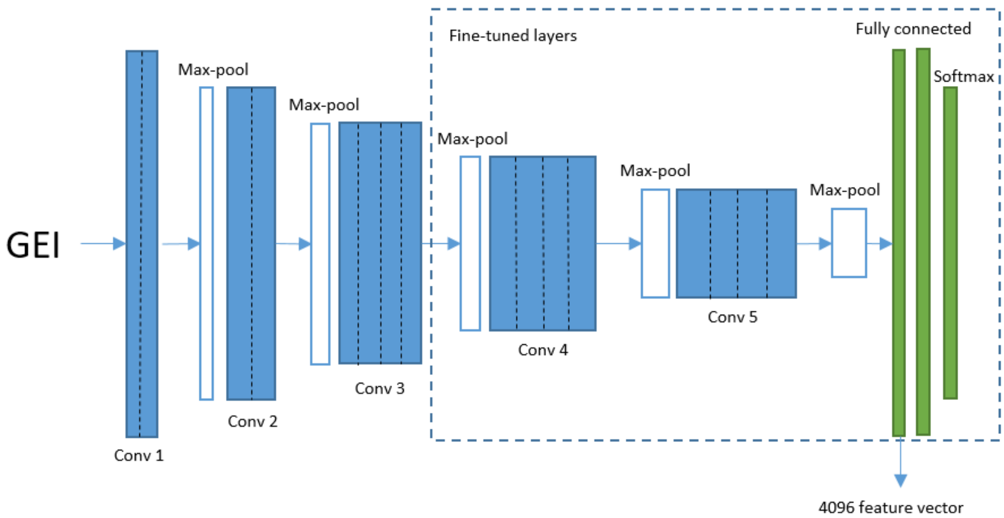


Figure 9.8: Feature extraction (VGG-19) architecture.

Although the complete VGG-19 can be used as a classifier, the output of the first fully connected layer is more effective as a feature vector [137] when the training data is small. Thus, the proposed system extracts the 4096-dimensional vector from the first fully connected layer as a feature vector. The resulting feature vector can then be used to perform pathology classification. Before performing the final classification step, the proposed system uses PCA and LDA for dimensionality reduction and data decorrelation. Classification of gait across different gait related pathologies could then be performed over an input GEI by comparing its features to the features registered in the database using k-NN.

9.4 PERFORMANCE EVALUATION

The performance evaluation section of the Thesis is divided into two halves. In the first half, the "hand-crafted" biomechanical feature-based system is evaluated using only the INIT dataset, as the features, such as shift in centre of gravity, torso orientation and amount of movement, cannot be reliably computed in the presence of silhouette segmentation errors present in DAI and DAI 2 datasets. The ability of each feature to differentiate between normal and impaired gait is evaluated using the two sample t-test. It is followed by the evaluation of the proposed system to check its accuracy in pathology classification. The second half evaluates the deep CNN-based pathology classification system. It first discusses the fine-tuning of the system using the INIT dataset. It then evaluates the system's performance using DAI and DAI 2 datasets.

9.4.1 Performance evaluation of the biomechanical feature-based system

The performance of the "hand-crafted" biomechanical feature-based system proposed in Section 9.2 is evaluated using the INIT dataset [118], which contains 10 individuals simulating 3 different gait related pathologies along with the normal gait, as described in Section 3.3.4. To evaluate the ability of each feature to differentiate between normal and impaired gait, a two sample t-test with unequal variances, with a significance level of 0.05 is conducted [140]. If the two feature samples are normally distributed, the t-test provides an assessment of the reliability of a given feature in being able to differentiate between normal and impaired gait. Thus, given two sample sets of a feature, drawn from the available normal gait sequences (NM) and one of the other sequence groups corresponding to gait with: (i) restricted full body movement (FB), (ii) restricted right leg movement (RL), and (iii) restricted left leg movement (LL), the test will either accept or reject a null hypothesis, which states that the two samples are drawn from the same population group. The decision is made using a p-value, which is the probability of finding the observed result, or an extreme one, when the null hypothesis of the t-test is true. If the p-value is above the significance level (in this case 0.05), the null hypothesis is accepted. The rejection of the null hypothesis suggests that the two sample sets are drawn from two different population groups. It indicates that the feature being tested is discriminative enough to differentiate between normal and impaired gait.

In Table 9.1, each entry presents the p-value of the two sample t-test. The bold entries in the table indicate the significant results. The first row of the table presents the results between the NM and FB groups. The results show significantly low p-values for almost all the computed features. The lowest values are observed for step length, amount of movement and torso orientation, suggesting that these features are more significant when differentiating between NM and FB gait. The low p-values are due to shorter step lengths, restricted body movement and a hunched posture. The hunched posture also causes a significant difference in the shift in the COG feature, represented by a low p-value in Table 9.1. The speed of the individuals in the NM group is also significantly different from the FB group. This is represented by low p-values for the speed, normalized step count and the foot flat ratio. The only feature that accepts the null hypothesis is the step length symmetry, as expected, indicating that when differentiating between NM and FB gait the step lengths of both the legs are similar, and so the symmetry feature cannot be used to identify this type of gait impairment. However, the step length symmetry and amount of movement symmetry features are significant when differentiating between NM and RL/LL gait, as indicated by their low p-values in Table 9.1. The p-value for the other features, such as step length, foot flat ratio, speed, normalized step count and shift in COG, are also low suggesting that they are significant enough to distinguish between the two groups. The features that accept the null hypothesis for RL/LL gait are the torso orientation and the amount of movement for the unrestricted side of the body. However, this is expected as the torso orientation feature is only effective in severe posture instability cases, such as hunchbacks, and the amount of movement of the unrestricted side is expected to be similar to NM group.

Table 9.1: Two sample t-test with unequal variances and significance level of 0.05 performed between normal and impaired gait.

	FB	**RL**	**LL**
SL Left	**1.56×10^{-21}**	**2.59×10^{-02}**	**4.01×10^{-09}**
SL Right	**1.05×10^{-23}**	**1.44×10^{-10}**	**3.88×10^{-02}**
SL Symmetry	1.74×10^{-01}	**5.29×10^{-10}**	**1.38×10^{-09}**
FFR Left	**4.48×10^{-07}**	**1.00×10^{-03}**	**4.89×10^{-04}**
FFR Right	**4.82×10^{-07}**	**3.25×10^{-04}**	1.08×10^{-01}
S	**7.02×10^{-16}**	**1.33×10^{-04}**	**1.48×10^{-04}**
NSC	**4.47×10^{-11}**	**2.50×10^{-05}**	**1.11×10^{-05}**
OT	**2.94×10^{-14}**	6.46×10^{-01}	8.75×10^{-01}
COG_{shift}	**6.87×10^{-03}**	**1.54×10^{-04}**	**4.97×10^{-02}**
AOM Left	**8.04×10^{-16}**	3.328×10^{-01}	**2.04×10^{-09}**
AOM Right	**1.98×10^{-19}**	**1.25×10^{-07}**	5.97×10^{-01}
AOM Symmetry	**2.46×10^{-3}**	**1.21×10^{-07}**	**1.94×10^{-09}**

To better illustrate the difference between the different types of gait, each entry in Table 9.2 presents the mean and standard deviation of each feature belonging to the respective group. As reported in Table 9.2, the FB gait is significantly slower than the NM gait, indicated by low speed, high NSC and a large fraction of time spent in foot flat during stance phase. The step lengths are also significantly shorter than in the NM group. However, there is no significant difference between the left and the right foot, as indicated by low step length symmetry values and the amount of movement symmetry values. In addition, the bending of torso and the shift in COG is significantly larger than for the NM gait – see Table 9.2. For the RL/LL groups, the restricted leg/side of the body is indicated by short step length and low amount of movement (entropy values) in Table 9.2. However, the gait is relatively fast as indicated by the higher speed and lower normalized step count values. Finally, it should be noted that the shift in COG feature is effective in differentiating between normal and all the three types of impaired gait, as reported in Table 9.2 , but it is not very precise in its measurement. The low precision in the measurement is caused by camera distortions, whose effect is severe, especially at the start and at the end of the gait sequence. Although its precision is not as good as that of other features, its ability to differentiate between different gait impairments allows the proposed system to classify the gait as being either normal or impaired.

Table 9.2: Mean and standard deviation of all the observed gait features belonging to different gait related pathologies.

	FB	RL	LL	NM
SL Left (pixels)	48.49±12.70	115.16±13.24	70.95±25.26	124.19±11.28
SL Right (pixels)	41.19±10.24	63.24±23.36	108.69±21.90	120.68±11.56
SL Symmetry (pixels)	7.28±6.18	51.92±18.61	45.90±17.20	5.137±3.08
FFR Left (%)	0.80±0.09	0.70±0.05	0.68±0.09	0.64±0.04
FFR Right (%)	0.75±0.09	0.66±0.05	0.67±0.06	0.60±0.06
S (pixels/sec)	37.04±8.86	63.11±15.31	64.77±13.44	81.41±11.37
NSC (steps/pixels)	0.025±0.005	0.013±0.002	0.013±0.001	0.010±0.000
OT (˚)	62.87±6.25	84.97±3.00	85.25±3.24	85.40±2.85
COG_{shift} (pixels)	12.39±5.93	4.65±2.25	6.40±2.85	8.08±2.84
AOM Left (entropy)	1.76±0.31	3.12±0.14	2.51±0.29	3.17±0.11
AOM Right (entropy)	1.58±0.25	2.35±0.42	3.01±0.24	3.10±0.11
AOM Symmetry (entropy)	0.17±0.12	0.77±0.39	0.55±0.21	0.06±0.05

The "hand-crafted" biomechanical feature-based system proposed in section 9.2 is further evaluated to check its ability to classify gait across different gait related pathologies. The classification accuracy of the proposed system is obtained using a fivefold cross-validation technique. This technique divides the data into five sets, where each set contains features from two different individuals. Thus, the training and testing set are mutually exclusive, with respect to the participating individuals. Next, the classification step is repeated 5 times such that each time, four sets are used for training and one set is used for testing the system. Finally, an average is computed to represent the classification accuracy of the system. The advantage of using fivefold cross-validation is that the variance of the resulting estimate is reduced, as the results do not depend on the partitioning the data. The classification accuracy of the proposed system is reported in Table 9.3, performing better than the state-of-the-art, being able to classify gait sequences as FB, RL, LL or NM with a correct classification accuracy of 98.8%. The proposed system can identify the left and the right leg, which allows a more complete characterization of gait impairments than what is possible with the current state-of-the-art. This can be concluded from the result reported in Table 9.3 that contains a comparison of the proposed system against the state-of-the-art markerless 2D vision-based systems, evaluated using the same fivefold cross-validation technique. The leg angle system presented [54], can be effective when there is enough separation between legs. However, even in NM group, it becomes difficult to distinguish between the two legs during mid stance and mid swing phases, while in the FB sequences there is no separation between the two legs during the entire gait cycle. The work presented in [28] uses a GEI along with SVM to perform classification of gait impairments. The use of the GEI allows the system to successfully differentiate between FB and NM groups, but there are significant misclassifications between the NM, RL and LL groups, reducing the overall classification accuracy to 75% – see Table 9.3. Even with a linear SVM, the proposed system performs better than the GEI-based system with a correct classification accuracy of 95.0%, following the same fivefold cross-validation technique. A second drawback of the GEI-based system [28] is that the GEIs used for the classification process do not provide any additional information about the gait impairments, while the proposed system provides measurable features that can be used to analyse the individuals' gait. Thus, it can be concluded that the biomechanical features used by the proposed system provide a good representation for gait impairment detection and classification.

Table 9.3: Classification accuracy (%) of the proposed "hand-crafted" biomechanical feature-based system and state-of-the-art systems in classifying between FB, RL, LL and NM gait.

Pathology classification system	Classification accuracy
Leg angles-based system [54]	72.5
GEI-based system [28]	75.0
Biomechanical features-based system (9.2)	**98.8**

The performance of the proposed system can also be analysed using the confusion matrix presented in Table 9.4. It shows that when using the proposed system only a single sequence is misclassified, which is due to the limited size of the available database, and results in a 5% penalty to the classification accuracy of the RL group. It should also be noted that the falsely classified sequence is a poor simulation of a RL gait impairment, as can be observed by the mean step length of the left and right legs, leading to its classification into the NM group.

Table 9.4: Confusion matrix (%) for the proposed biomechanical features-based system.

		Ground truth			
		FB	**RL**	**LL**	**NM**
Prediction	**FB**	100	0	0	0
	RL	0	95	0	5
	LL	0	0	100	0
	NM	0	0	0	100

9.4.2 Performance evaluation of the deep CNN-based system

The deep CNN-based pathology classification system proposed in Section 9.3 is evaluated using the INIT [118], DAI [56] and DAI 2 [28] datasets, discussed in Section 3.3.4. However, unlike the hand-crafted system presented in Section 9.2, this system uses the INIT dataset only for fine-tuning the weights of the VGG-19 network. The other two datasets can then be used to train the classifier and evaluate the performance of the proposed system. The DAI dataset is used to evaluate the ability of the proposed system to classify gait as either normal or impaired, and the DAI 2 dataset is used to perform classification of gait across different gait related pathologies.

Since the size of the three available datasets is relatively small, it is decided to take 60% of the largest dataset for fine-tuning the CNN weights, and use the remaining gait sequences of that dataset for validation. Evaluation can then be performed on the other two datasets, using the previously adopted CNN weights, to test the generalization ability of the obtained model.

The INIT dataset is thus selected to fine-tune VGG-19, as it includes 20 video sequences for each of the four pathology groups. Since each sequence captures at least two gait cycles, a GEI for every gait cycle is generated, increasing the total number of GEIs to 160. As mentioned above, the dataset is split into a training set, with 60% of the sequences, and a validation set, with the remaining 40%. To further increase the size of the training set and to make the network robust to minor changes such as flips, scale changes and translations, data augmentation is performed on the training dataset. Data augmentation allows the system to account for situations not foreseen in the original training set, such as walking in the opposite direction of that available in the training sequences. Each GEI in the training set is thus augmented using small shifts, shear, zoom, as well as horizontal flipping, resulting in 480 GEIs.

The pre-trained VGG-19 model has been optimized to perform classification across the 1000 ImageNet image groups. In the proposed system the classification is performed using LDA, while the output from the first fully connected layer is used as a feature vector. However, to fine-tune the network for pathology classification the final softmax layer of the VGG-19 network is replaced by a different softmax layer that performs classification only across the four INIT dataset groups considered. Fine-tuning is done using backpropagation on the validation set. It considers a learning rate of 0.001, as further increasing the learning rate can lead to convergence problems. The batch size is set to 34 to use optimally the available graphic card memory size, and the number of training epochs is set to 150, with early stopping to prevent overfitting. The remaining parameters, such as dropout regularisation and loss function, maintain their default settings.

For the fine-tuning of the VGG-19, several alternatives are considered, notably hanging the set of layers whose weights are adjusted when running the backpropagation optimization. In a first experiment, only the fully connected layers (FC) are re-trained. In the following experiments, the FC layers along with one or more convolutional layers (CONV) are re-trained. This process is repeated while re-training an extra CONV layer at each experiment. The classification accuracy over the training and the validation sets is reported in Table 9.5. The final column reports the results on the validation set after replacing the softmax layer with the proposed LDA classifier. During each experiment, the LDA classifier is trained using features extracted by the fine-tuned VGG-19 over the INIT training set.

As reported in Table 9.5, the best results are obtained by re-training the fully connected and the convolutional layers 4 and 5 for 38 epochs. Further training other layers reduces the accuracy in the validation set, indicating overfitting of the model. Thus, this configuration was selected for feature extraction. It can also be concluded from the results in Table 9.5 that the fine-tuned VGG-19 is better suited for feature extraction rather than classification of gait pathologies, as the LDA classifier outperforms the softmax classifier included in the VGG-19 architecture. The proposed system achieves a classification accuracy of 89% on the INIT dataset through fine-tuning. Using the same evaluation protocol, the "hand-crafted" biomechanical features-based system presented in Section 9.2 achieves a classification accuracy of 95%. However, its performance degrades significantly in the presence of segmentation errors, present in DAI and DAI 2 datasets.

Table 9.5: VGG-19 configurations and the corresponding classification accuracy (%).

Trained VGG-19 blocks	**Classification accuracy**		
	Training	**Validation (softmax classifier)**	**Proposed system (LDA classifier)**
FC layers	100	80	79
FC + CONV 5 layers	100	81	85
FC + CONV 5, 4 layers	**100**	**84**	**89**
FC + CONV 5, 4, 3 layers	100	81	82

Once VGG-19 is fine-tuned, the proposed system can use it to obtain features for the classifier. Since the VGG-19 is fine-tuned using the INIT dataset, the proposed system can be evaluated using the DAI and DAI 2 datasets. These two datasets respectively contain 2 and 5 different types of gait pathologies, as discussed in Section 3.3.4. Thus, the LDA classifier must be trained separately on the two datasets. The training and the testing sets are obtained using a fivefold cross-validation technique, which divides the data into five mutually exclusive sets of individuals. The process is repeated 5 times, such that each time four sets are used for training and fifth sets is used for testing the system. Finally, an average is computed to represent the classification accuracy of the system.

Before evaluating the classification accuracy of the proposed system, the quality of features obtained from several deep CNNs such as ResNet50, Xception, InceptionV3, VGG-16 and VGG-19, trained on ImageNet, are considered. The resulting features are not specifically fine-tuned for classification of gait across different pathologies. However, even without fine-tuning, VGG-16 and VGG-19 achieve a classification accuracy of above 90% across both datasets. As reported in Table 9.6, the accuracy of the system improves when VGG-16 is replaced with VGG-19 for the feature extraction step. Since the deeper VGG-19 network performs better than VGG-16, it is selected for the fine-tuning process. It should also be noted that although VGG-19 is fine-tuned for the INIT Dataset, the resulting model improves the classification accuracy across the other two datasets. This is significant because the INIT dataset is captured in a LABCOM studio [118], which produces perfectly segmented silhouettes. The other two datasets, however, are captured in less constrained environments, where the segmentation of silhouettes is far from perfect, often missing parts of the walking person’s silhouette. Thus, it can be concluded that the proposed fine-tuning scheme generalizes well across datasets, even in the presence of silhouette segmentation errors, which affect the performance of the systems based on biomechanical features – see Section 9.2.

Table 9.6 also reports results for the state-of-the-art systems. Among them, only the GEI system [28] operates on both the datasets, as this system also uses the GEI, making it robust to silhouette segmentation errors. The GEI system [28] performs well with the binary classification problem of the DAI Gait dataset, but its performance degrades significantly when the number of pathologies increases. Finally, the leg angle-based system [54] is very effective in performing binary classification over the DAI dataset but, as reported in [28], it cannot be used to perform pathology specific classification.

Table 9.6: Classification accuracy (%) of the system proposed and the state-of-the-art.

Classification Systems	**Classification accuracy**		
	DAI	**DAI 2**	**Mean**
Leg angles-based system [54]	**100**	-	-
GEI-based system [28]	97	74	86
Proposed system + VGG-16	90	92	91
Proposed system + VGG-19	92	94	93
Proposed system + VGG-19 (fine-tuned)	97	**95**	**96**

Although the proposed system does not have the best performance in all situations across both datasets, it performs consistently well under different conditions, even when its feature extraction module has been trained on a dataset different from the ones considered for testing. It also provides the best results in the presence of silhouettes with segmentation errors, which can be a challenging task for the current state-of-the-art 2D vision-based systems. This is reported in Table 9.6, where the proposed system performs better than the state-of-the-art with a mean classification accuracy of 96%. Thus, it can be concluded that the proposed system is best suited for classification of such gait pathologies.

It should also be noted that the proposed system could distinguish between different gait pathologies with a high level of certainty, as reported in the confusion matrix in Table 9.7. Only the gait affected by hemiplegia presents a classification accuracy lower than 90%, due to its similarities with gait affected by diplegia and neuropathy. This is a significant improvement over the GEI system [28], which fails in classifying diplegia, with a classification accuracy of only 40 %. It also performs poorly in the classification of gait affected by hemiplegia and neuropathy. Hence, the proposed system can be considered a step forward, when compared to the current state-of-the-art 2D vision-based systems.

Table 9.7: Confusion matrix (%) for the system proposed in Section 9.2 operating on DAI 2 dataset.

		Ground truth				
		Diplegia	**Hemiplegia**	**Neuropathy**	**Parkinson**	**Normal**
Prediction	**Diplegia**	98	2	0	0	0
	Hemiplegia	8	87	5	0	0
	Neuropathy	0	6	94	0	0
	Parkinson	2	0	0	98	0
	Normal	0	6	0	0	94

Part IV: Conclusion

10 FINAL REMARKS

10.1 INTRODUCTION

This Thesis highlights the use of gait in unconstrained environments for the purpose of biometric recognition and pathology classification. To address the limitations of the existing gait-based biometric recognition systems, such as change in the viewpoint of the camera and change in the appearance of the individuals being observed, this improved solutions are proposed in this Thesis. New gait representations that improve the recognition accuracy of gait-based biometric recognition systems, capturing only the identity information while preserving health-related and other private data, are presented in this Thesis. New areas in gait-based biometric recognition, such as the use of shadow silhouettes to obtain features for the recognition systems, are also explored. The Thesis also presents new systems for medical diagnosis using gait. A novel system, proposed in this Thesis, can estimate the IC and the TO events as well as temporal gait features with a high level of accuracy, using a single 2D video camera. Two novel 2D vision-based systems for pathology classification are also presented in this Thesis. The first system explores the use of biomechanical features to perform gait-based pathology classification. The second system uses GEI along with deep CNN to classify gait across different gait related pathologies. This chapter presents a summary of the contributions followed by a discussion about the possible future directions.

10.2 SUMMARY OF CONTRIBUTIONS

The work presented in this Thesis focuses on the use of gait acquired from a 2D video camera to perform biometric recognition and pathology classification in unconstrained environments.

10.2.1 Gait-based biometric recognition

The gait-based biometric recognition systems can be affected by several factors, notably related to the individuals being observed, the camera characteristics, the light source and the environment. The problems caused by these factors or the combination thereof can be perceived as changes in the viewpoint of the camera or changes in the appearance of the individuals being observed. The current literature lacks an encompassing discussion of such factors. Therefore, a novel taxonomy discussing the factors affecting the gait-based biometric recognition systems is presented in this Thesis.

Two novel gait representations are presented in this Thesis that improve the recognition accuracy of the gait-based biometric recognition system. The GDV proposed in Section 4.2 explores the dissimilarity space, resulting in a gait representation that captures only the identity information of the individuals being observed, unlike other gait representations, such as the GEI, which can reveal information about the individuals' health, gender or age. The representation is obtained by computing the Euclidean distance between an input GEI and a representation set called the prototype. Using this representation allows achieving a recognition accuracy of 99.6% when tested on CASIA B dataset, which is equivalent to the best state-of-the-art 2D vision-based systems.

The gait representation proposed in Section 4.3 called the SEGI can be obtained for the individuals registered in the database by applying RPCA to the GEIs belonging to each individual.

Given an input GEI, its SEGI representations are generated with respect each individual registered in the database using their GEIs. Recognition can then be performed by associating the identity of the individual to the SEGI with the smallest Euclidean norm/Euclidean distance. The proposed SEGI representation is evaluated using CASIA B dataset, which results in a recognition accuracy of 99.2%. The result is equivalent to the best state-of-the-art systems that employ better performing classifiers. The simplicity of the matching module used in evaluating the proposed gait representations highlights their effectiveness in gait-based biometric recognition.

Novel systems that address the problems faced by the gait-based biometric recognition systems operating in unconstrained environments, such as the change in the viewpoint of the camera and change in the appearance of the individuals being observed are also presented in this Thesis. To address the problem of change in the viewpoint of the camera a novel system is proposed in Section 5.2 that applies a PHash function to the leg region of an input GEI, to obtain a hash value that represents the general structure of the leg region. The hash value can then be matched with the database, using k-NN with Hamming distance as measure, to detect the viewpoint of the camera. The proposed system performs significantly better than the state-of-the-art viewpoint detection systems, with an average accuracy of 97.0% when evaluated on the CASIA B dataset. However, a limitation of the proposed system is that it can detect only those viewpoints that are registered in the database. Thus, a second system is presented in Section 5.3, which performs viewpoint detection without the use of a database. The system captures the evolution of the feet position of the individuals over time using the bottom part of a GTI contour. The feet positions represent the dominant walking trajectory of the individuals. Thus, the viewpoint of the camera can be estimated using the direction of the walking trajectory. The proposed system also achieves an average viewpoint detection accuracy of 97.0% on CASIA B dataset, but without any need for previous training.

The two proposed viewpoint detection systems operate using the leg region or the feet positions of the observed individuals. Thus, the presence of shadows under the feet can affect their performance. To tackle this limitation, a third novel system is presented in Section 5.4, which performs viewpoint detection by fitting a line through the feet positions such that the direction of the line represents the viewpoint of the camera. The system identifies the feet position by analysing the intensity values of the GTI.

The problem of appearance change is addressed in this Thesis by presenting a novel system in Section 6.2 that selects only the unaltered sections of a GEI to perform gait-based biometric recognition. The proposed system decomposes a GEI into 11 sections and compares them to the average GEI image, obtained by averaging the available GEIs in the database, to identify the unaltered sections. Each unaltered section is matched with the corresponding sections registered in the database. Recognition is performed by a majority voting decision among the unaltered sections. The proposed system performs significantly better than the state-of-the-art systems with an average recognition accuracy of 94.0% on CASIA B dataset.

The Thesis also explores the use of shadow silhouette for gait-based biometric recognition, as under certain conditions the shadow cast by the individuals appears similar to the body silhouettes and can be an alternative source of gait features. However, the shadow silhouettes may contain distortions and deformations caused by the camera perspective and parameters that can hamper the performance of the system. The two novel systems presented in this Thesis rectify the shadow silhouettes, so that they can be successfully used to perform gait-based biometric recognition. The first novel system presented in Section 7.2 uses TILT to obtain a transformation matrix that rectifies the input silhouettes into a canonical viewpoint. GEIs obtained from the rectified silhouettes can then be used to perform recognition. When evaluated using the KY IR Shadow dataset, the proposed system achieves a recognition accuracy of 94.0%. The recognition accuracy can be further improved to 97.0% using the 4-point correspondence system presented in Section 7.3. Being an optimisation problem TILT cannot be fully controlled. The novel 4-point correspondence system uses the head and feet positions of the shadow silhouettes, and their projections into the canonical viewpoint, to obtain the transformation matrix. The rectification of the shadow silhouettes using this transformation matrix is significantly better than the TILT system. Thus, the GEIs obtained from the rectified silhouettes result in improved recognition results.

The shadow silhouettes used by gait-based biometric recognition systems must be sharp and appear similar to the body of the individuals being observed. The usability of the shadow silhouettes can be checked using the novel system proposed in Section 7.4. The system generates a histogram of intensity ratios computed between the foreground containing the individuals and their shadow and the static background. The usable sharp shadows and the other diffused shadows are both represented with two distinct plots. Sharp shadows are saturated with low intensity ratio values, while diffused shadows contain a significant amount of both low and high intensity ratio values. Thus, the shadow type can be identified by analysing the peaks in these histograms.

To evaluate the performance of the systems presented in Chapter 7, the IST shadow gait dataset and the IST shadow type dataset are presented in Section 7.5. The IST shadow gait dataset is used to evaluate the 4-point correspondence system presented in Section 7.3. The results suggest that under certain conditions the use of rectified shadow silhouettes can even provide better performance than using the body silhouettes. The IST shadow type dataset is used to evaluate the shadow type identification system, presented in Section 7.4.

10.2.2 Gait-based pathology classification

The use of gait analysis for medical diagnosis is explored in Chapters 8 and 9 of this Thesis. A novel system is presented in Section 8.2 that estimates temporal gait features by detecting the occurrences of the IC and the TO during a gait cycle. The system detects the IC and the TO by analysing the overlap between the foot flat position and the position of the feet during a gait cycle. The proposed system estimates the IC and the TO events with an accuracy of 99.0% on the CASIA A dataset. The system performs equally well on the CASIA B dataset, which contains viewpoint changes. The system is also evaluated using the IST-KUL gait dataset presented in Section 8.3. This dataset allows to compare the results of the proposed system, which operates without markers and using a single 2D camera, with those of a motion capture system, considered the gold standard for this type of analysis. The obtained results indicate ‘good’ to ‘excellent’ correlation between the two systems. The correlation suggests that the proposed system estimates the temporal gait features in a very reliable way using a simple and inexpensive 2D camera.

Chapter 9 of the Thesis proposes two novel systems that perform pathology classification using video sequences acquired from a 2D video camera. The first system, presented in Section 9.2, uses biomechanical features to classify gait across different gait related pathologies. The features acquired by the proposed system can be classified into two groups. The first group is related to the feet of the individuals. They include features such as step length, normalized step count, speed and the fraction of foot flat during a stance phase. The proposed system can distinguish among the features obtained from the left and the right foot, thus allowing to estimate gait symmetry. The second group of features are related to the entire body of the individuals. It include features such as the amount of movement while walking, torso orientation and the shift in the COG with respect to its COS. Features such as the amount of movement are computed separately for each foot, allowing to compute a symmetry score. Apart from detecting the left-right symmetry, the proposed system also detects posture instabilities, using torso orientation and the shift in COG, while other features such as the normalized step count and speed are used to detect the deviation from the normal gait. Using an SVM classifier the proposed system achieves a classification accuracy of almost 100.0% across four different of pathologies, using the INIT dataset.

The second system, presented in Section 9.3, is capable of classifying gait related pathologies caused by neurological or systemic disorders such as diplegia, hemiplegia, neuropathy and Parkinson’s diseases with an accuracy of 95%. It operates even in situations where most state-of the-art systems fail, such as in the presence of poorly segmented silhouettes. The proposed system tackles this problem by using a gait representation called the GEI. To further improve the classification accuracy, the proposed system obtains the best features from the GEI using the VGG-19 deep CNN, fine-tuned on INIT dataset. The results indicate that VGG-19 fine-tuned for the classification of gait pathologies performs significantly better than deep CNNs, such as VGG-16 or VGG-19, trained on ImageNet, while also generalizing well to other datasets, such as the DAI and the DAI 2 datasets.

10.3 Future research directions

The work presented in this Thesis highlights the effectiveness of analysing gait sequences captured using a 2D camera for biometric recognition and pathology classification systems operating in unconstrained environments. Several problems faced by such systems are addressed in this Thesis, but there is scope for further improvement and development of these systems. This section summarises some possible future research directions.

10.3.1 Gait-based biometric recognition

Three novel systems to detect the viewpoint of the camera are presented in Chapter 5 of this Thesis. These systems operate under the assumption that the dominant walking direction of the individuals remains fixed along the observed gait sequence. If the individuals change their walking direction in between, the observed features will undergo a viewpoint change, as illustrated in Figure 10.1. The problem of viewpoint change along curved trajectories is significantly more difficult, as a single gait cycle contains features belonging to several different viewpoints, which can lead to poor recognition results. Model-based systems can perform well along curved trajectories because of the availability of additional information about the camera parameters and the scene – see Section 3.2.1. Such information is unavailable for 2D vision-based systems.

A possible solution to the problem can be the rectification of the silhouettes into a canonical viewpoint. Unwrapping a GTI along a curved surface using techniques such as RPCA [141] can result in a transformation matrix that can be used to rectify silhouettes into the canonical viewpoint. RPCA is currently used to rectify textures along curved surfaces without any additional information about the camera parameters or the scene. Thus, their effectiveness in rectifying silhouettes can be evaluated as future work. Another possible solution can include fitting a 3D skeletal model onto a 2D image, for instance by using deep CNNs, such as 3D pose [142]. Such systems can capture the 3D skeletal model of the individuals without any camera calibrations, depth information or information about the scene, making them suitable for usage in unconstrained environments.

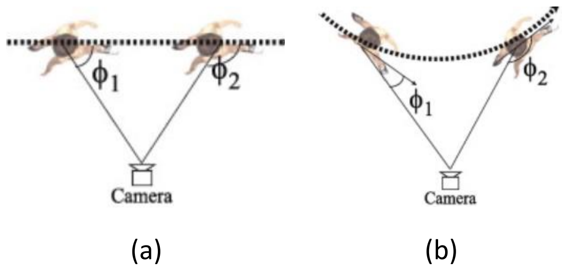


Figure 10.1: Illustration of an individual (a) walking along a straight line, (b) walking along a curved trajectory

A second possible future direction can involve the use of shadow silhouettes for gait-based biometric recognition. The novel systems proposed in Chapter 7 of this Thesis address the problems faced by the state-of-the-art due to camera perspective and parameters. However, the proposed systems operate under the assumption that the position of the light source illuminating the scene remains unchanged throughout the day. As illustrated in Figure 10.2, the shadow cast by the individuals change with changes in the position of the light source. From early to mid-morning, the low position of the sun in the sky casts a long, side view shadow. From mid-morning to noon the shadow turns shorter as the sun moves to an overhead position. Noon to mid-evening the shadow starts turning longer with the sun now illuminating the other side of the individual. Finally, mid to late evening the sun casts a long shadow illuminating the side again. Thus, the shadow cast from mid-morning to mid-evening can be expected to provide poor gait features, due to the overhead position of the light source. The change in the quality of the shadow silhouettes along the day must be studied to improve the performance of the recognition systems that use the shadow silhouettes.

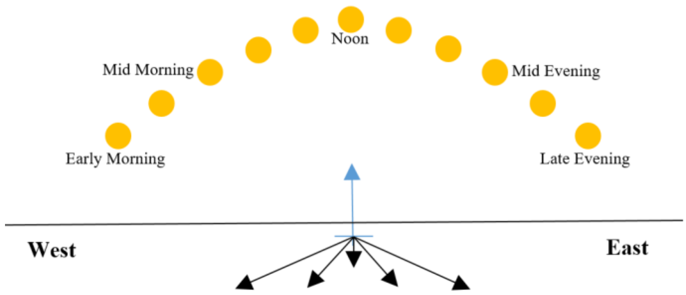


Figure 10.2: Shadow cast throughout the day.

10.3.2 Gait-based pathology classification

To perform classification of gait across different gait related pathologies or to acquire accurate gait features suitable for clinical evaluations, the systems presented in this Thesis rely on the use of silhouettes. The quality of the silhouettes depends on several factors, which cannot be controlled in unconstrained environments, as discussed in Section 2.4. A possible solution to such problems can be obtaining gait features directly from the given video sequence. Deep CNNs, such as 3D pose [142], can fit a 3D skeletal model onto a 2D image without any additional information about the camera or the scene. It obtains the skeletal model using a sequence of CNNs that repeatedly produce and improve on the belief map for the location of each key point in the image. The belief map is generated by learning the image dependent spatial models of the relationship between key points. Thus, such systems can detect the key points even under occlusions, as illustrated in Figure 10.3. Although the use of deep CNNs appears promising, the information used by them is still 2D. Thus, its performance against a motion capture system still needs to be evaluated in the future.

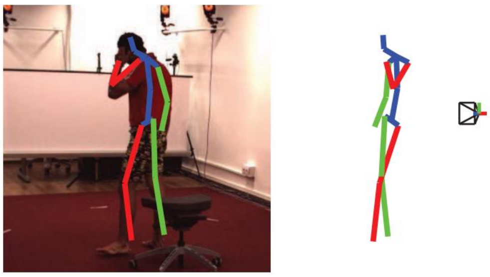


Figure 10.3: Output of the 3D pose system [142].

Novel systems to perform classification of gait across different gait related pathologies are presented in Chapter 9 of this Thesis. The evaluation of those systems can be further improved by using a large-scale dataset. The largest dataset currently available is the INIT dataset, which contains only 4 different types of gait, with only 10 individuals recorded per gait type. The different types of gait themselves do not correspond to any specific gait pathology, as discussed in Section 3.3.4.. To obtain significant results, a new dataset must be captured containing large number of individuals actually suffering from different types of gait pathologies. Since finding individuals suffering from such pathologies can be difficult, a possible future direction can include the capture of a large-scale dataset containing different type of simulated gait pathologies.

Another possible alternative can involve the acquisition of a large-scale dataset where individuals simulate different stages of different gait pathologies. As illustrated inFigure 10.4, the deterioration of gait due to Parkinson's diseases is gradual and can be classified into 5 stages. Availability of such datasets will allow the development of systems that can estimate the severity of the illness using a 2D camera, in a daily life setting.

Figure 10.4: Deterioration of gait due to Parkinson's diseases.

Finally, a possible future direction involves the combination of all the work presented in this Thesis, i.e., performing gait based biometric recognition on individuals suffering from various gait related pathologies. Individuals in hospitals or elderly homes suffering from various disorders are very unlikely to cooperate with the recognition systems due to factors, such as deterioration of the memory or frailty of the body. Under such conditions automatic access control to their rooms or other facilities can turn extremely difficult. Since gait analysis can be performed without the cooperation of the individuals, and can reveal their identity and pathology, it is well suited to be used for feature acquisition in automatic access control. However, the change in the quality of the features representing the individuals' identity due to deterioration of the gait is yet to be studied.

REFERENCES

[1] T. Verlekar, L. Soares and P. Correia, “Automatic classification of gait impairments using a markerless 2D video-based system,” *MDPI Sensors,* vol. 18, no. 9, p. 2743, 2018.

[2] A. Muro-De-La-Herran, B. Garcia-Zapirain and A. Mendez-Zorrilla, “Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications,” *Sensors,* vol. 14, no. 2, pp. 3362-3394, 2014.

[3] Y. Makihara, D. Matovski, S. Nixon, J. Carter and Y. Yagi, “Gait recognition: databases, representations, and applications,” *Wiley Encyclopedia of Electrical and Electronics Engineering,* 2015.

[4] Y. Wahab and N. A. Bakar, “Gait analysis measurement for sport application based on ultrasonic system,” in *IEEE 15th International Symposium on Consumer Electronics (ISCE)*, 2011.

[5] L. T. Kozlowski and J. E. Cutting, “Recognizing the sex of a walker from a dynamic point-light display,” *Perception & psychophysics,* vol. 21, no. 6, pp. 575-580, 1977.

[6] N. F. Troje, “ Decomposing biological motion: A framework for analysis and synthesis of human gait patterns,” *Journal of vision,* vol. 2, no. 5, p. 2, 2002.

[7] “https://jogger.co.uk/how-to-improve-your-running-technique-sports-physiotherapist-chris-bramah/”.

[8] Aristotle, On the gait of animals, 350 B.C..

[9] G. A. Borelli, On the motion of animals (De Motu Animalum), 1680.

[10] E. Heinrich, W. Eduard and E. F. W. Weber, Die Mechanik Der Menschlichen Gerverzeuge, 1836.

[11] E. J. Marey, La Methode Graphique, Paris, 1885.

[12] W. Braüne and O. Fischer, Der Gang des Menschen (1895-1904), Berlin: Springer-Verlag, 1987.

[13] J. Saunders, V. Inman and H. Eberhart, “The major determinants in normal and pathological gait,” *Journal of Bone and Joint Surgery.,* vol. 35, no. A, pp. 543-558, 1953.

[14] J. Perry, Gait analysis: normal and pathological function, Thorofare, New Jersey: SLACK Incorporated, 1992.

[15] D. Sutherland, R. Olshen, E. Biden and W. M., The development of mature walking, Cambridge University Press., 1988.

[16] M. P. Murray, G. B. Spurr, S. B. Sepic, M. Gardner G. and L. A. Mollinger, “Treadmill vs. floor walking: kinematics, electromyogram, and heart rate,” *Journal of Applied Physiology,* vol. 59, pp. 87-91, 1985.

[17] J. E. Cutting and L. T. Kozlowski, “Recognizing friends by their walk: Gait perception without familiarity cues,” *Bulletin of the Psychonomic Society,* vol. 9, no. 5, p. 353–356, 1977.

[18] J. Vanrenterghem, D. Gormley, M. Robinson and A. Lees., “Solutions for representing the whole-body centre of mass in side cutting manoeuvres based on data that is typically available for lower limb kinematics,” *Gait & posture,* vol. 31, no. 4, 2010.

[19] I. Bouchrika, A survey of using biometrics for smart visual surveillance: Gait recognition, Springer, 2018, pp. 3-23.

[20] “https://www.bayometric.com/identification-verification-segmented-identification/”.

[21] K. Jain A. and A. Kumar, “Biometric recognition: an overview,” in *Second generation biometrics: the ethical, legal and social context*, 2012.

[22] A. Stoica, “Towards recognition of humans and their behaviors from space and airborne platforms: Extracting the information in the dynamics of human shadows,” in *Bio-inspired, Learning and Intelligent Systems for Security*, Edinburgh, 2008.

[23] “http://robotics.ait.kyushu-u.ac.jp/yumi/publications.html”.

[24] J. A. DeLisa, Gait analysis in the science of rehabilitation (Vol. 2), Diane Publishing, 1998.

[25] “https://www.pinterest.com/pin/336995984590985656/”.

[26] K. E. Webster, J. E. Wittwer and J. A. Feller, “Validity of the GAITRite walkway system for the measurement of averaged and individual step parameters of gait,” *Gait Posture,* vol. 22, no. 3, p. 17–21, 2005.

[27] J.-S. Wang, C.-W. Lin, Y.-T. C. Yang and Y.-J. Ho, “Walking pattern classification and walking distance estimation algorithms using gait phase information,” *IEEE Transactions on Biomedical Engineering,* vol. 59, no. 10, pp. 2884-2892, 2012.

[28] M. Nieto-Hidalgo and J. M. García-Chamizo, “Classification of pathologies using a vision based feature extraction,” in *2017*, International Conference on Ubiquitous Computing and Ambient Intelligence.

[29] T. Verlekar, H. Vroey, K. Claeys, H. Hallez, L. Soares and P. Correia., “robust estimation and validation of biomedical gait indicators using 2D video,” *Computer methods and programs in biomedicine*, vol. 175, 2019.

[30] T. Verlekar, L. Soares and P. Correia, “Gait recognition in the wild using shadow silhouettes,” *Image and Vision Computing,* vol. 76, 2018.

[31] T. Verlekar, P. Correia and L. Soares, “View-invariant gait recognition system using a gait energy image decomposition method,” *IET Biometrics,* vol. 6, no. 4, 2017.

[32] T. Verlekar, P. Correia and L. Soares., “Using transfer learning for classification of gait pathologies,” in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM*, Madrid, Spain, 2018.

[33] T. Verlekar, P. Correia and L. Soares, “Gait recognition using normalized shadows,” in *European Signal Processing Conference, EUSIPCO*, Kos Island, Greece, 2017.

[34] T. Verlekar, P. Correia and L. Soares, “Sparse error gait image: A new representation for gait recognition,” in *International Workshop on Biometrics and Forensics, IWBF*, Coventry, United Kingdom, 2017.

[35] T. Verlekar, P. Correia and L. Soares, “View-invariant gait recognition exploiting spatio-temporal information and a dissimilarity metric,” in *International Conference. Of the Biometrics Special Interest Group, BIOSIG*, Darmstadt, Germany,, 2016.

[36] T. Verlekar, L. Soares and P. Correia, “Shadow type identification for gait recognition using shadows,” in *23th Portuguese Conference on Pattern Recognition, RECPAD*, Lisbon, Portugal, 2017.

[37] S. Shultz, P. Houglum and D. Perrin, “Examination of musculoskeletal injuries with web resource,” *Human Kinetics,* 2015.

[38] J. Kaye, M. Swift and S. Bell, “The clinical orthopedic assessment guide,” *Human Kinetics,* 2008.

[39] W. Pirker and R. Katzenschlager, “Gait disorders in adults and the elderly: a clinical guide,” *Wien Klin Wochenschr,* vol. 129, p. 81–95, 2017.

[40] A. M. Howell, T. Kobayashi, H. Hayes, K. Foreman and S. Bamberg, “Kinetic gait analysis using a low-cost insole,” *IEEE Transactions on Biomedical Engineering,* vol. 60, no. 12, pp. 3284-3290, 2013.

[41] G. Barry, D. McGrath, R. O’Neill, K. J. O’Donovan, A. Burns and B. Caulfield, “An adaptive gyroscope-based algorithm for temporal gait analysis,” *Medical & biological engineering & computing,* vol. 48, no. 12, pp. 1251-1260, 2010.

[42] C. Wentink, H. Schut, C. Prinsen, S. Rietman and H. Veltink, “Detection of the onset of gait initiation using kinematic sensors and EMG in transfemoral amputees,” *Gait & posture,* vol. 39, no. 1, pp. 391-396, 2014.

[43] “https://www.bestclouds.us/custom-cycling-shoes-boulder/”.

[44] “https://www.biomech-solutions.com/en/analisis-movimiento-tecnologia-inercial-xsens.html”.

[45] “http://www.shimmersensing.com/products/consensys-ecg-development-kits-update”.

[46] L. Middleton, A. Buss, A. Bazin and M. Nixon, “A floor sensor system for gait recognition,” in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, 2005.

[47] A. A. Chaaraoui, R. P. L. José and F.-R. Francisco, “Abnormal gait detection with RGB-D devices using joint motion history features,” in *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2015.

[48] G. Zhao, G. Liu, H. Li and M. Pietikainen, “3D gait recognition using multiple cameras,” in *7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[49] "https://www.tekscan.com/strideway".

[50] "http://www.hsc.umn.edu/Facility.html".

[51] "https://www.researchgate.net/publication/284171406_An_Open_Source_Framework_to_Manage_Kinect_on_the_Web".

[52] W. A. Stuberg, V. L. Colerick, D. J. Blanke and W. Bruce, "Comparison of a clinical gait analysis method using videography and temporal-distance measures with 16-mm cinematography," *Physical therapy,* vol. 68, no. 8, pp. 1221-1225, 1988.

[53] Z. Z. Heijden. and F. V. Der, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters,* vol. 27, no. 7, pp. 773-780, 2006.

[54] M. Nieto-Hidalgo, F. Ferrández-Pastor, R. Valdivieso-Sarabia, J. Mora-Pascual and J. García-Chamizo, "A vision based proposal for classification of normal and abnormal gait using RGB camera," *Journal of Biomedical Informatics,* vol. 63, no. 3, pp. 82 - 89, 2016.

[55] M. M. Serrano, Y.-P. Chen, A. Howard and P. A. Vela, "Automated feet detection for clinical gait assessment," in *IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, 2016.

[56] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual and J. M. García-Chamizo, "Vision based extraction of dynamic gait features focused on feet movement using RGB camera," *Ambient Intelligence for Health, Springer, Cham,* pp. 155-166, 2015.

[57] J. Fazenda, P. Correia and S. D., "Using gait to recognize people," in *The International Conference on Computer as a Tool, EUROCON*, Belgrade, 2005.

[58] D. S. Choudhury and T. Tardi, "Robust view invariant multiscale gait recognition," *Pattern Recognition,* vol. 48, no. 3, pp. 798-811, 2015.

[59] J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich and D. Schulte, "Embodiment of sadness and depression: gait patterns associated with dysphoric mood," *Psychosomatic medicine,* vol. 71, no. 5, pp. 580-587, 2009.

[60] R. Zisserman and H. Andrew, "Multiple view geometry in computer vision.," Cambridge University Press., 2003, p. 155–157.

[61] Y. Guan, Y. Sun, C.-T. Li and M. Tistarelli, "Human gait identification from extremely low-quality videos: an enhanced classifier ensemble method," *IET biometrics,* vol. 3, no. 2, pp. 84-93, 2014.

[62] N. Jia, V. Sanchez, C. T. Li and H. Mansour, "On reducing the effect of silhouette quality on individual gait recognition: a feature fusion approach," in *IEEE International Conference of the Biometrics Special Interest Group (BIOSIG),*, 2015.

[63] T.-N. Nguyen, H.-H. Huynh and M. Jean, "Extracting silhouette-based characteristics for human gait analysis using one camera," in *Fifth Symposium on Information and Communication Technology. ACM*, 2014.

[64] Y. Iwashita, R. Kurazume and A. Stoica, “Gait identification using invisible shadows: robustness to appearance changes,” in *Fifth International Conference on Emerging Security Technologies (EST)*, Alcala de Henares, 2014.

[65] S. Sarkar, J. Phillips, L. Zongyi, R. Isidro, G. Patrick and K. Bowyer, “The humanid gait challenge problem: Data sets, performance, and analysis,” *IEEE transactions on pattern analysis and machine intelligence,* vol. 27, no. 2, pp. 162-177, 2005.

[66] P. Wohlberg and B. Rodriguez, “Fast principal component pursuit via alternating minimization,” in *IEEE International Conference on Image Processing*, 2013.

[67] H. Ji, C. Liu, Z. Shen and Y. Xu, “Robust video denoising using low rank matrix completion,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

[68] R. Sohrab and B. Mahdi, “Human gait recognition using body measures and joint angles,” *International journal of scientific knowledge computing and information technology,* vol. 6, no. 4, p. 2305, 2015.

[69] W. Yufei, S. Jiande, L. Jing and Z. Dong, “Gait recognition based on 3D skeleton joints captured by kinect,” in *IEEE International Conference on Image Processing (ICIP)*, 2016.

[70] G. Zhao, G. Liu, H. Li and P. M., “3D gait recognition using multiple cameras,” in *7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[71] Y. Iwashita, R. Kurazume, R. Baba and K. Ogawara, “Method for gait-based biometric identification robust to changes in observation angle,” in *26th International Conference of Image and Vision Computing*, Auckland, 2011.

[72] G. Zhao, G. Liu, H. Li and M. Pietikainen, “3D gait recognition using multiple cameras,” in *7th International Conference on Automatic Face and Gesture Recognition*, Southampton, 2006.

[73] D. Muramatsu, A. Shiraishi, Y. Makihara and Y. Yagi, “Arbitrary view transformation model for gait person authentication.,” in *Biometrics: Theory, Applications and Systems (BTAS), IEEE Fifth International Conference on*, 2012.

[74] Y. Iwashita, R. Kurazume and K. Ogawara, “Identification of people walking along curved trajectories,” *Pattern Recognition Letters,* vol. 48, pp. 60-69, 2014.

[75] F. Jeana, A. B. Albub and R. Bergevina, “Towards view-invariant gait modeling: Computing view-normalized body,” *Pattern Recognition,* vol. 42, no. 11, pp. 2936-2949, 2009.

[76] M. Goffredo, I. Bouchrika, J. Carter and M. Nixon, “Self-calibrating view-invariant gait biometrics,” *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 40, no. 4, pp. 997 - 1008, 2010.

[77] I. Bouchrika, J. Carter and M. Nixon, “Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras,” *Multimedia Tools and Applications,* vol. 75, no. 2, pp. 1201-1221, 2016.

[78] A. Kale, A. Roy, K. Chowdhury and R. Chellappa, “Towards a view invariant gait recognition algorithm,” in *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003.

[79] J. Man and B. Bir, "Individual recognition using gait energy image," *IEEE transactions on pattern analysis and machine intelligence,* vol. 28, no. 2, pp. 316-322, 2006.

[80] Z. Lv, X. Xing, K. Wang and D. Guan, "Class energy image analysis for video sensor-based gait recognition: a review," *Sensors,* vol. 15, pp. 932-964, 2015.

[81] K. Bashir, X. Tao and G. Shaogang, "Gait recognition using gait entropy image," in *3rd International Conference on Crime Detection and Prevention (ICDP).*, 2009.

[82] W. Kusakunniran, Q. Wu, H. Li and Z. Jian, "Multiple views gait recognition using view transformation model based on optimized gait energy image," in *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, Kyoto, 2009.

[83] N. Liu and Y. P. Tan, "View invariant gait recognition," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, 2010.

[84] W. Kusakunniran, Q. Wu, H. Li and J. Zhang, "Support vector regression for multi-view gait recognition based on local motion feature selection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[85] W. Kusakunniran, Q. Wu, H. Li and J. Zhang, "Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron," *Pattern Recognition Letters,* vol. 7, no. 33, pp. 882-889, 2012.

[86] T. Connie, M. Goh and A. Teoh, "A grassmannian approach to address view change problem in gait recognition," *IEEE transactions on cybernetics,* 2016.

[87] T. Connie, M. Goh and A. Teoh, "Multi-view gait recognition using a doubly-kernel approach on the Grassmann manifold," *Neurocomputing,* vol. 216, pp. 534-542., 2016.

[88] W. Kusakunniran, Q. Wu, H. Li and Z. Jian, "A new view-invariant feature for cross-view gait recognition," *Information Forensics and Security, IEEE Transactions,* vol. 8, no. 10, pp. 1642-1653, 2013.

[89] B. Khalid, X. Tao and G. Shaogang, "Cross view gait recognition using correlation strength," in *British Machine Vision Conference*, London, 2010.

[90] Y. Guan, C.-T. Li and Y. Hu, "An adaptive system for gait recognition in multi-view environments," in *14th ACM Multimedia and Security Workshop*, Coventry, UK, 2012.

[91] Y. Liang, C.-T. Li, Y. Guan and Y. Hu, "Gait recognition based on the golden ratio," *Journal on Image and Video Processing,* no. 22, 2016.

[92] Y. Pratheepan, J. V. Condell and G. Prasad, "P RW GEI: Poisson random walk based gait recognition," in *IEEE 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2011.

[93] N. Jain, M. Jeevan and M. Hanmandlu, "Gait recognition based on gait pal and pal entropy image," in *20th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2013.

[94] K. Uchino, R. K. and Y. Iwashita, "Gait-based person identification robust to changes in appearance," *Sensors,* vol. 13, no. 6, pp. 7884-7901, 2013.

[95] K. Bashir, X. Tao and G. Shaogang, “Gait recognition without subject cooperation,” *Pattern Recognition Letters,* vol. 31, no. 13, pp. 2052-2060, 2010.

[96] Y. Iwashita, A. Stoica and R. Kurazume, “Person identification using shadow analysis,” in *British Machine Vision Conference, BMVC 2010*, Aberystwyth, 2010.

[97] Y. Iwashita and A. Stoica, “Gait recognition using shadow analysis,” in *Symposium on Bio-inspired Learning and Intelligent Systems for Security*, Edinburgh, 2009.

[98] Y. Iwashita, A. Stoica and R. Kurazume, “People identification using shadow dynamics,” in *17th IEEE International Conference on Image Processing (ICIP)*, Hong Kong, 2010.

[99] Y. Iwashita, R. Kurazume and A. Stoica, “Finding people by their shadows: Aerial surveillance using body biometrics extracted from ground video,” in *Third International Conference on Emerging Security Technologies (EST)*, 2012.

[100] Y. Iwashita, A. Stoica and R. Kurazume, “Gait identification using shadow biometrics,” *Pattern Recognition Letters,* vol. 33, no. 16, p. 2148–2155, 2012.

[101] Y. Iwashita, K. Uchino, R. Kurazume and A. Stoica, “Gait identification from invisible shadows,” in *Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX*, Baltimore, Maryland, 2012.

[102] M. Shinzaki, Y. Iwashita, R. Kurazume and K. Ogawara, “Gait-based person identification method using shadow biometrics for robustness to changes in the walking direction,” in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, 2015.

[103] R. Gross and J. Shi, “Tech. report CMU-RI-TR-01-18,” Robotics Institute, Carnegie Mellon, 2001.

[104] R. Tanawongsuwan and A. Bobick, “Tech. report GIT-GVU-03-01,” Georgia Institute of Technology, 2003.

[105] T. Chalidabhongse, V. Kruger and R. Chellappa, “Tech. report,” Univ. of Maryland, 2001.

[106] W. Liang, T. Tieniu, N. Huazhong and H. Weiming, “Silhoutte analysis based gait recognition for human identification,” *IEEE trans Pattern Analysis and Machine Intelligence(PAMI),* vol. 25, no. 12, pp. 1505-1518, 2003.

[107] J. Shutler, M. Grant, N. M.S. and C. J.N., “On a large sequence-based human gait database,” in *4th International Conference on Recent Advances in Soft Computing*, 2002.

[108] S. Sarkar, J. Phillips, Z. Liu, I. Robledo, P. Grother and K. Bowyer, “The Human ID gait challenge problem: data sets, performance, and analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 2, pp. 162-177, 2005.

[109] Y. Shiqi, T. Daoliang and T. Tieniu, “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in *The 18’th International Conference on Pattern Recognition (ICPR)*, Hong Kong, China., 2006.

[110] a. Iwama, M. Okumura, Y. Makihara and Y. Yagi, “the OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition,” *IEEE Trans. on Information Forensics and Security,* vol. 7, no. 5, pp. 1511-1521, 2012.

[111] “http://ayunext.com/casestudy/gait-abnormalities/”.

[112] M. Gabel, G.-B. Ran, R. Erin and S. Assaf, “Full body gait analysis with Kinect,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012.

[113] R. Z. L. Hu, A. Hartfiel, J. Tung, A. Fakih, J. Hoey and P. Poupart, “3D Pose tracking of walker users' lower limb with a structured-light camera on a moving platform,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011.

[114] E. E. Stone and S. Marjorie, “Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing,” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC,* 2011.

[115] A. Leu, R. Durrant and G. Axel, “A robust markerless vision-based human gait analysis system,” *6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI),* pp. 415-420, 2011.

[116] K. W. Lin, S. T. Wang, P. C. Chung and C. F. Yang, “A new view-calibrated approach for abnormal gait detection,” *In Advances in Intelligent Systems and Applications,* vol. 2, pp. 521-529, 2013.

[117] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual and J. M. García-Chamizo, “Vision based gait analysis for frontal view gait sequences using RGB camera,” in *10th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI*, 2016.

[118] J. Ortells, M. T. Herrero-Ezquerro and R. A. Mollineda, “Vision-based gait impairment analysis for aided diagnosis,” *Medical & biological engineering & computing,* pp. 1-12, 2018.

[119] R. Krishnan, M. Sivarathinabala and S. Abirami, “Abnormal gait detection using lean and ramp angle features,” *Computational Intelligence in Data Mining,* vol. 1, pp. 325-335, 2016.

[120] T.-N. Nguyen, H.-H. Huynh and J. Meunier, “Extracting silhouette-based characteristics for human gait analysis using one camera,” in *Proceedings of the Fifth Symposium on Information and Communication Technology. ACM*, 2014.

[121] C. Bauckhage, J. K. Tsotsos and F. E. Bunn, “Automatic detection of abnormal gait,” *Image and Vision Computing,* vol. 27, no. 1, pp. 108-115, 2009.

[122] R. Duin and E. Pękalska, “The dissimilarity space: Bridging structural and statistical pattern recognition,” *Pattern Recognition Letters,* vol. 33, no. 7, pp. 826 - 832, 2012.

[123] E. Candès, X. Li, Y. Ma and J. Wright, “Robust principal component analysis?,” *Journal of the ACM (JACM),* vol. 58, no. 3, p. 11, 2011.

[124] H. Chaubey, H. Madasu and S. Vasikarla, “Enhanced view invariant gait recognition using feature level fusion,” in *In Applied Imagery Pattern Recognition Workshop (AIPR)*, Washington, DC, 2014.

[125] C. Zauner, “Implementation and benchmarking of perceptual image hash functions,” Upper Austria University of Applied Sciences, Hagenberg, 2010.

[126] M. Fischler and R. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM,* vol. 24, no. 6, pp. 381-395, 1981.

[127] Z. Zhengdong, X. Liang, A. Ganesh and Y. Ma, “TILT: Transform invariant low-rank textures,” *International journal of computer vision,* vol. 99, no. 1, pp. 1-24, 2012.

[128] M. Robinson and J. Vanrenterghem, “An evaluation of anatomical and functional knee axis definition in the context of side-cutting,” *Journal of biomechanics,* vol. 45, no. 11, pp. 1941-1946, 2012.

[129] D. Slijepcevic, M. Zeppelzauer, A. M. Gorgas, C. Schwab, M. Schüller, A. Baca and B. Horsak, “ Automatic Classification of Functional Gait Disorders,” *IEEE Journal of Biomedical and Health Informatics,* 2017.

[130] K. Wang, K. Delbaere, M. A. Brodie, N. H. Lovell, L. Kark, S. R. Lord and S. J. Redmond, “Differences between gait on stairs and flat surfaces in relation to fall risk and future falls,” *IEEE journal of biomedical and health informatics,* vol. 21, no. 6, pp. 1479-1486, 2017.

[131] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 9, no. 1, pp. 62-66, 1979.

[132] J. Fleiss and J. Cohen, “The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability,” *Educational and psychological measurement,* vol. 33, no. 3, pp. 613-619, 1973.

[133] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556,* 2014.

[134] S. Choudhury and T. Tjahjadi, “Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors,” *Pattern Recognition,* vol. 45, no. 9, pp. 3414-3426, 2012.

[135] C. Shannon, “A mathematical theory of communication,” *ACM SIGMOBILE mobile computing and communications review,* vol. 5, no. 1, pp. 3-55, 2001.

[136] E. Allwein, R. Schapire and Y. Singer, “Reducing multiclass to binary: A unifying approach for margin classifiers,” *Journal of machine learning research,* vol. 1, pp. 113-141, 2000.

[137] C. Li, X. Min, S. Sun, W. Lin and Z. Tang, “Deepgait: a learning deep convolutional representation for view-invariant gait recognition using joint bayesian,” *Applied Sciences,* vol. 7, no. 3, p. 210, 2017.

[138] R. Liao, C. Cao, E. Garcia, S. Yu and Y. Huang, “Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations,” in *Chinese Conference on Biometric Recognition*, 2017.

[139] M. Hon and N. Khan, “Towards Alzheimer's disease classification through transfer learning,” in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017.

[140] C. Chatfield, the analysis of time series: An introduction, Chapman & Hall, New York, 1989.

[141] Z. Zhang, X. Liang and Y. Ma, “Unwrapping low-rank textures on generalized cylindrical surfaces,” in *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[142] M. Lin, L. Lin, X. Liang, K. Wang and H. Cheng, “Recurrent 3d pose sequence machines,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.