Human Identification from Video Using Advanced Gait Recognition Techniques

Sabesan Sivapalan
BSc Eng (Hons)

A Thesis Submitted in Fulfilment
of the Requirements for the Degree of

Doctor of Philosophy

Queensland University of Technology

Image and Video Research Laboratory
Science and Engineering Faculty

September 2014

Keywords

human identification, biometrics, gait recognition, segmentation, gait energy image, principal component analysis, multiple discriminant analysis, 3D reconstruction, gait energy volume, Microsoft Kinect®, frontal depth gait, local directional pattern, local binary pattern, histogram of oriented gradients, compressive sensing, sparse representation, back-filled gait energy image, view invariant gait recognition, pose estimation

Abstract

Research into biometrics is an ongoing and open research challenge with the aim of achieving robust human identification in a visual surveillance environment. Compared to other biometrics, gait has attracted significant attention in recent years due to the unique advantages that other biometrics may not offer. Most importantly, it can be used with video feeds obtained at a distance without alerting the subject and with low-resolution video. Interest in gait has increased significantly, due to the promising recognition results obtained from research in this area under constrained environments. Recent research is more focused on improving recognition results in realistic environments, where it is necessary to address the effects of changes in view, resolution and fluctuation of gait patterns, due to carrying goods or different footwear or clothes.

In this dissertation, real world solutions that can handle the gait challenging conditions are proposed to assist security and forensic applications. Initially, 3D space is considered through reconstruction from multiple views using synchronised multi-view silhouette images. A fast 3D ellipsoidal-based gait recognition algorithm, using these reconstructed 3D voxel models, is proposed. This approach directly solves the limitations of view dependency and self-occlusion in existing ellipse fitting model-based approaches. Features derived from the ellipsoid parameters are modelled using a Fourier representation to retain the temporal dynamic pattern for classification. Improved recognition performances are achieved over its 2D counterpart on the CMU Motion of Body (CMU MoBo) database.

Gait energy images (GEIs) and their variants form the basis of many recent appearance-based gait recognition systems. The GEI combines good recognition performance with a simple implementation, though it suffers problems inherent to appearance-based approaches, such as being highly view dependent. The concept of the GEI has been extended to 3D, to create what is called as the *gait energy volume*, or *GEV*. A basic GEV implementation is tested on the CMU MoBo database, showing improvements over both the GEI baseline and a fused multi-view GEI approach.

Having a multi-view camera set-up to enable multi-view GEI & GEV-based approaches can be impractical under many applications. This is particularly the case in concise spaces where most biometric-authentication systems are installed and operated. An alternative to acquiring this 3D data would be to use a depth sensing device. Frontal-based depth has the advantage of being able to capture essentially all characteristics of a person's gait from a single viewpoint without the issue of self-occlusion. A frontal viewpoint also makes it possible to easily integrate into biometric portals to assist security services. The efficacy of the GEV approach is explored on the partial frontal volumes that have been synthesised from multi-view data. With the promising results obtained, an in-house frontal depth gait database (DGD) is developed and state-of-the-art performance is demonstrated using a new proposed frontal GEV feature.

One of the main limitations in appearance-based methods is that extracted features need to be robust towards unwanted variations in the image, whether they be due to lighting or pose. Different types of patch-based gradient feature extraction methods are explored to minimise the effects caused by these these variations. Extending existing popular feature extraction methods, a histogram of oriented gradients (HOG) and local directional pattern (LDP), we propose a novel technique, histogram of weighted local directions (HWLD). In addition to this feature optimisation, a Sparse representation-based classifier (SRC) is also proposed for the robust classification in a gait recognition context. Evaluations on the popular gait databases show the highest recognition rate for the proposed HWLD method. In addition, the HWLD is extended to 3D, which is demonstrated using the GEV feature on the DGD dataset, observing further improvements in performance.

Finally, a novel direction has been proposed for the gait recognition research by proposing a new capture-modality independent, appearance-based feature, the Back-filled Gait Energy Image (BGEI). It can be constructed from both frontal depth images, as well as the more commonly used side-view silhouettes, allowing the feature to be applied across these two differing capturing systems, using the same enrolled database. The results demonstrate that the BGEI can effectively be used to identify subjects through their gait across these two differing input devices, achieving a rank-1 match rate of 100% in the conducted experiments.

The proposed solutions in this thesis contribute to improve gait recognition performance in various practical scenarios that further enable the adoption of gait recognition into real world security and forensic applications.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

BGEI	Backfilled Gait Energy Image
CASIA	The gait database from Chinese Academy of Sciences
CMS	Cumulative Match Score
CMU MoBo	Carnegie Melon University Motion of Body Database
DGD	Depth Gait Database
DOF	Degree Of Freedom
FRGC	Face Recognition Grand Challenge
FAR	Face Alarm Rate
GEI	Gait Energy Image
GEV	Gait Energy Volume
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HWLD	Histogram of Weighted Local Directions
HOG	Histogram of Oriented Gradients
ISD	Independent Score for similarity Distance
KPCA	Kernel Principal Component Analysis
LBP	Local Binary Pattern
LDP	Local Directional Pattern
MDA	Multiple Discriminant Analysis
MEI	Motion Energy Image
MHI	Motion History Image
MRF	Markov Random Field
OULP	Osaka University Large Population gait database
PCA	Principal Component Analysis
ROC	Receiver Operating Curve
SRC	Sparse Representation Classification
SSD	Single Score for similarity Distance
SVD	Singular Value Decomposition

List of Publications

Conference papers

- **S. Sivapalan**, D. Chen, S. Denman, S. Sridharan, C. Fookes, "3D Ellipsoid Fitting for Multi-view Gait Recognition," in *Proceedings, 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Klagenfurt, Austria, 2011, pp. 355–360.

- **S. Sivapalan**, D. Chen, S. Denman, S. Sridharan, C. Fookes, "Gait Energy Volumes and Frontal Gait Recognition using Depth Images," in *Proceedings, International Joint Conference on Biometrics (IJCB)*, Washington DC, USA, 2011, pp. 1-6.

- **S. Sivapalan**, R. Rana, D. Chen, S. Sridharan, S. Denman, C. Fookes, "Compressive Sensing for Gait Recognition," in *Proceedings, IEEE International Conference on Digital Image Computing: Techniques and Applications DICTA*, Noosa, Australia, 2011, pp. 567–571.

- **S. Sivapalan**, D. Chen, S. Sridharan, S. Denman, C. Fookes, "The Backfilled GEI – A Cross-capture Modality Gait Feature for Frontal and Side-view Gait Recognition," in *Proceedings, 2012 International Conference on Digital Image Computing: Techniques and Applications DICTA*, Freemantle, Australia, 2012, pp. 1–8.

- **S. Sivapalan**, D. Chen, S. Denman, S. Sridharan, C. Fookes, "Local Directional Pattern Analysis for Gait Recognition," in *Proceedings, IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Portland, OR, USA, 2013, pp. 125–130.

Journal papers

- **S. Sivapalan**, D. Chen, S. Sridharan, S. Denman, C. Fookes, “Is Appearance-based Gait recognition Really Gait Recognition? - A Survey,” in Computer Vision and Image Understanding (2014) (under review).
- D. Chen, **S. Sivapalan**, S. Denman, S. Sridharan, C. Fookes, “Optimised Directional Kernel-based Inter-frame Registration for Robust Gait Recognition,”in Computer vision and image understanding (2014) (under review).

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

QUT Verified Signature

Signature: ____________________

Date: 29/09/2014

Acknowledgements

First of all, I would like to thank Professor Sridha Sridharan for giving me the opportunity to undertake this research program. Thanks also to my other supervisors, Associate Professor Clinton Fookes and Dr Simon Denman for their support and guidance throughout my candidature. Their encouragement and support has been most appreciated. I would especially like to thank Dr Daniel Chen for helping me to overcome technical challenges encountered during my candidature. Furthermore, I would like to extend my thanks to everyone else in the SAIVT lab for their friendship and assistance. I would also like to thank Ms Diane Kolomeitz for providing copyediting and proofreading services, according to the guidelines laid out in the university-endorsed national guidelines for editing research theses.

I would like to thank to my college peers and family members for giving me a much needed break at times, in the form of a social life. Special gratitude is reserved for my parents who deserve every success of mine as their own. Finally, I would like to thank my ever-loving wife, Priatharsny, for her support and understanding on the occasions when matters related to my research weren't proceeding well.

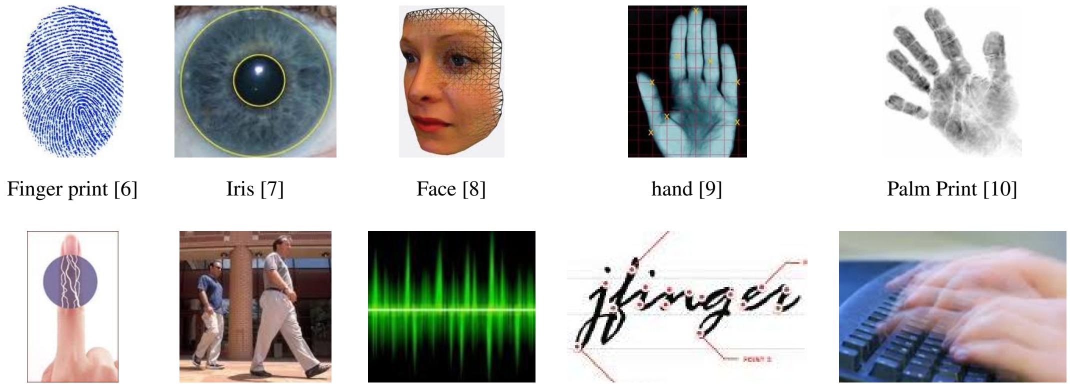
Chapter 1

Introduction

1.1 Motivation and Overview

Human recognition has been an active research area since the need for reliable user authentication techniques has increased, from providing access control in technologically advanced applications such as inter-networking, to many common and widespread applications across surveillance, security, forensics, physical and logical access control and border control for an entire country [2–4]. In addition to the traditional authentication techniques such as pins and password, biometrics have now gained widespread acceptance to provide legitimate authentication of identity for an individual [5], particularly in computer-vision based applications.

Biometrics is the science of establishing the identity of an individual, based on inherent physical and/or behavioural traits associated with a person [16]. A typical



Finger print [6] Iris [7] Face [8] hand [9] Palm Print [10]

Finger Vein [11] Gait [12] Voice [13] Signature [14] Keyboard stroke [15]

Figure 1.1: Examples of biometrics.

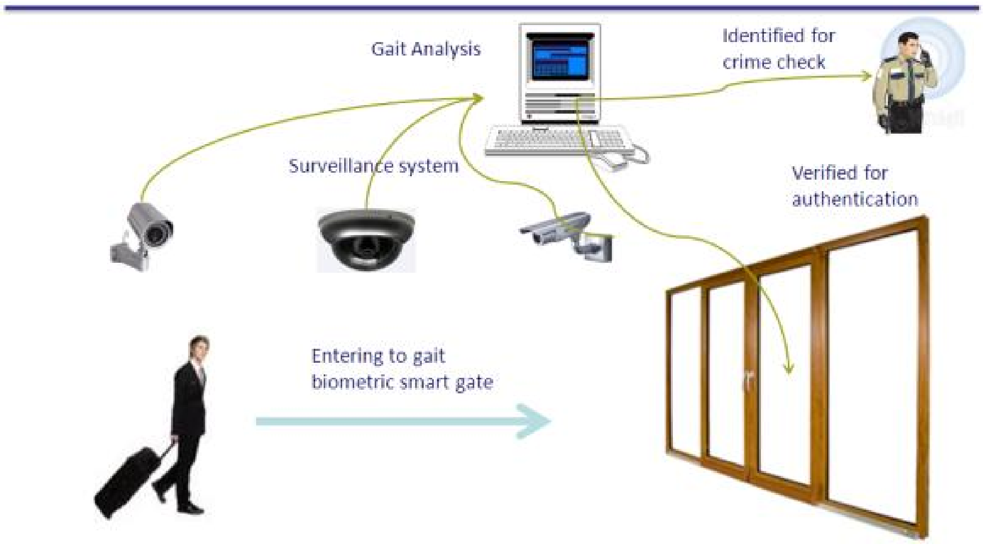

Gait Analysis
Identified for crime check
Surveillance system
Verified for authentication
Entering to gait biometric smart gate

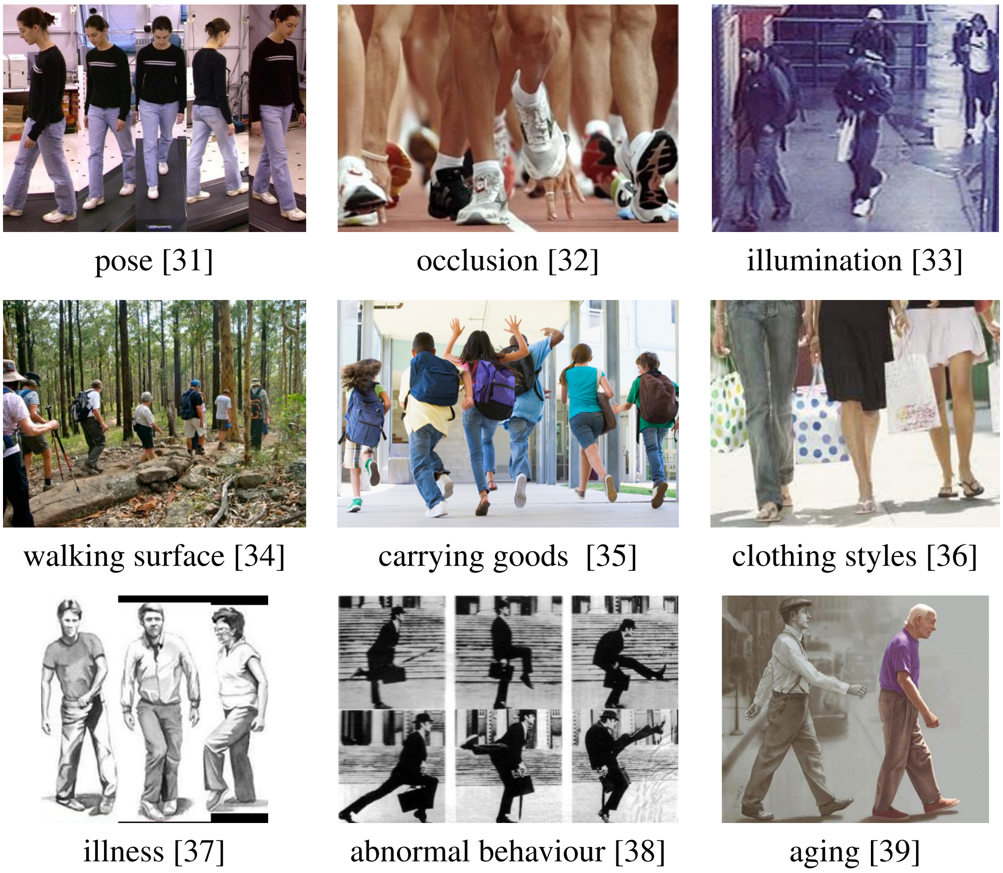
Figure 1.2: Application of conceptual gait recognition system.

biometric-based authentication system operates by extracting the required features and comparing those features with the already registered biometric samples in the database, to validate the claimed identity or identify the person. Fingerprint, iris, face, ear, hand geometry, palm print, finger vein geometry, gait, voice, signature, keyboard stroke pattern *etc*. (Figure 1.1) are popular biometrics used in various systems [2, 17–28]. **Gait**, as a biometric, is attractive due to its ability to operate using low resolution imagery, and can be acquired at a distance without alerting the subject [29].

The intentional objective of a gait-based authentication system is illustrated in Figure 1.2. Let us assume the CEO of a company walks steadily towards the main entrance of his office, situated at the end of a well-lit corridor, equipped with several cameras. As he gets close to the entrance, his gait is verified, the door opens automatically, and the intelligent system that manages the building, welcomes him with a friendly voice. When an unauthorised person gets close to the doors, his gait is not verified and his authentication is blocked. Further, if his gait is identified as a person on a criminal watch list, before he becomes a possible threat, all security alerts are activated. Today, there is no practical system that can support the above authentication scenario. However, recent research on gait-based authentication and recognition provides evidence that such a system is becoming realistic [30].

Gait can be defined as a *coordinated, cyclic combination of movements that results in human locomotion* [40]. The movements are coordinated in the sense that they must occur with a specific temporal pattern. From a biomechanics standpoint, human ambulation consists of synchronised, integrated movements of hundreds of muscles and joints of the body. Although these movements follow the same basic bipedal pattern for all humans, they seem to vary from one individual to another in certain details, such as their relative timing and magnitudes, i.e. the kinematics of gait [41]. The movements in a gait pattern repeat as a walker cycles between steps with alternating feet. It is both the coordinated and cyclic nature of the motion and the static human shapes. Gait is hence presumed to be unique to each individual, simply because it is determined by the totality of their musculo-skeletal structure. Features such as the stride length, bounce, rhythm, speed, swagger and physical lengths of human parts all contribute to the ability of a human to distinguish gait [42].

In automated gait recognition, gait features can be obtained using surveillance video footage, without interacting with, or alerting the subject, even in a low resolution environment. Other biometrics may not provide the required accuracy under these conditions [43]. Furthermore, gait represents behavioural and physical nature as a combined representation, which is difficult to impersonate without hampering the



pose [31]

occlusion [32]

illumination [33]

walking surface [34]

carrying goods [35]

clothing styles [36]

illness [37]

abnormal behaviour [38]

aging [39]

Figure 1.3: Common challenges in gait recognition.

subject's movement. Gait also can be used when other biometrics are obscured. For example, criminal intent might motivate concealment of the face, but it is difficult to conceal and/or disguise motion, as this generally impedes movement [44]. Because of all these potential advantages, gait has attracted significant attention in recent years in human identification applications.

Vision-based automatic gait recognition is not without its own problems, however. It is limited in being able to recognise an individual with fully occluded clothes or who behaves abnormally due to illness or ageing. However, in these situations, an individual can be easily distinguished for abnormality and can be processed through a manual authentication system [45]. Other than these unusual events, external factors also can affect the walking style of an individual, such as clothing styles, footwear, carrying things and the walking surface [45]. In addition, it also needs to address common vision-based issues such as occlusion, changes in pose/view or illumination changes. Examples illustrating these gait challenging factors are shown in Figure 1.3.

Recent gait recognition algorithms have shown 100% correct classification rate under controlled conditions [46]. However, these results significantly degrade in real-world data with the changes in the above gait-challenging conditions. Current gait recognition research is tackling the open research challenges that prevent the use of gait in security surveillance applications in public places, such as airports, shopping malls, and other transportation hubs. In addition to the authentication applications, gait analysis can also be used in medical applications for abnormality detection and in the field of animation, etc. [47].

A gait recognition system typically consists of the fundamental tasks illustrated in Figure 1.4. Most of the vision-based gait recognition systems are initiated by extracting the human silhouette (image pixels attributed to the shape of an individual) from video footages in order to extract the spatio-temporal behaviour of a moving person. These extracted silhouettes then need to be pre-processed for optimised registration, with proper alignment and normalisation. Next, different computer vision and machine learning techniques are used to extract and model various discriminative features. Following this, samples are generated and stored to form a database/ dictionary in the registration process. During the authentication process, a test sample is formed as above, and compared with the formed dictionary for identification or to validate a person's claim.

Within the elements in the gait recognition framework, computing the robust gait feature is mostly focussed on in recent research, as it is the key factor on which all

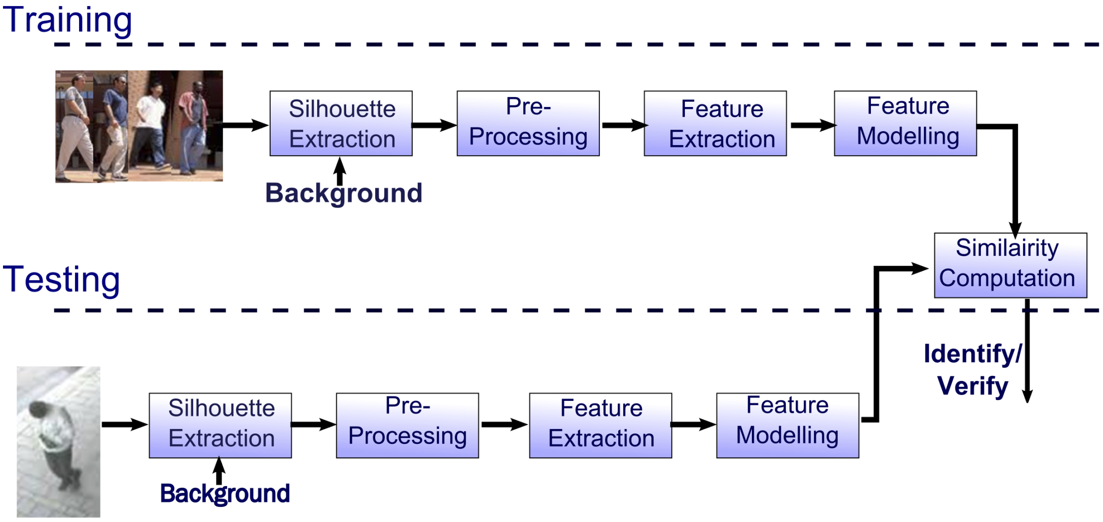


Figure 1.4: The gait recognition framework.

other elements depend. Gait feature extraction algorithms are generally classified as appearance-based or model-based. Model-based techniques gather gait dynamics directly, by modelling the underlying kinematics of human motion, whereas appearance-based methods try to establish correspondence between successive frames, based upon the implicit notion of what is being observed [48].

Both approaches attempt to address the main challenges in gait recognition to be invariant with changes in viewing angle (the angle of the subject to the camera), clothing, walking surface, the subject's walking style due to carrying objects, shoes or an injury. Model-based techniques are comparatively less susceptible to these changes, though recognition performances are poor, due to inaccurate model fitting that is very sensitive to the quality of image data. However, by utilising the static and dynamic temporal features, appearance-based methods provide better performance. Again, the recognition performance is very sensitive to the mentioned challenges.

This PhD thesis explores both of these methods, to develop innovative solutions to real world automatic gait recognition systems, utilising various signal processing and machine learning techniques.

1.2 Aims and Scope

Real world automatic gait recognition struggles to perform better as it is affected by gait challenging conditions. The solution to address these challenges will be to develop an appropriate feature, utilising the available source of information. Each real world environment has different sources of information with different levels of limitation. For example, public hubs or surveillance environments may have multiple cameras that have a wide view angle to capture an individual, while a security portal, where people line up for authentication is a concise place, where a single front camera is possible. These different sources of data and limitation need to be addressed separately for their optimised performances. However, an independent platform also needs to be there for cross-validation.

From the literature, it has been noted that recognition performance of model-based techniques are significantly poor compared to appearance-based techniques in similar conditions. In addition, model-based techniques are computationally expensive due to the complex matching and searching that they require for model-fitting. Even though these limitations in model-based methods motivate the appearance-based methods to follow, they also need to be addressed for performance degradation due to appearance changes.

Overall, the research in this thesis aims to improve the performance of gait recognition systems by producing an original contribution in the following areas: (a) gait recognition using multi-view data; (b) frontal gait recognition; (c) gait recognition in a cross-capture modality platform; (d) pre-processing and feature optimisation to handle appearance-changes; and (e) robust classifier for better recognition. These scenarios and research areas will be described as follows.

Improvements in the gait recognition performance in a multi-camera environment

Gait recognition performance is very sensitive to subject walking direction, particularly in 2D appearance-based methods. However, in most security authentication scenarios in public places such as airports, it is likely there are multiple views of the particular person. This research aims to investigate ways to incorporate this multi-view data to solve the problem of view-dependency in 2D gait recognition algorithms. This investigation results in the 3D reconstruction of the subject and gait feature extraction in 3D. Further investigation needs to be completed to reduce the segmentation and reconstruction noises in 3D to improve the performance. Working in 3D aims to address the following research questions:

1. Do features from 3D reconstructed voxel-volumes improve the gait recognition performance when a multi-view camera setup is available?

2. How can improved model-based and appearance-based gait recognition be achieved in 3D for better recognition performance?

To address these questions, appropriate gait features using model-based and appearance-based methods need to be computed and compared with well performing algorithms in single views and combined multiple-views.

Improvements in frontal gait recognition

Though the full 3D reconstruction can solve the view dependency issues, it is limited to the availability of a multiple-camera set-up. It is impractical, particularity in concise spaces such as a narrow corridor or security portals. However, frontal-view has its own merits in these scenarios; it doesn't need more space since a walking person approaches the camera. Biometrics, such as iris and face used in these scenarios, are implemented to capture in frontal. Therefore, it is required that gait also needs to be acquired in frontal to integrate with them as a distance biometric. However, frontal 2D images fail to gather all the gait dynamics and therefore frontal gait in 2D struggles to perform better recognition. To address this issue, the following research questions will be explored,

1. Can the loss of gait dynamics in 2D front-view be better represented using depth images?

2. How can the volume reconstruction and 3D analysis be explored in the frontal depth domain?

3. Is it possible to achieve adequate recognition in frontal-view by only considering the reconstructed volumes of frontal-surface and discarding the back-leg information?

4. What will be an effect of losing back leg gait information on distinguishing an individual? Furthermore, can acceptable recognition performance be achieved by considering only the frontal surface of the full 3D volume to avoid the requirement of multi-view camera setup?

5. How can frontal gait recognition be integrated with walk-through biometrics portal (portals with multiple biometric-capturing sensors to enable automatic human identification as people walk)?

To answer these questions, this research aims to explore frontal-gait recognition using depth-images and the methods to incorporate gait recognition into portal-based security applications.

Gait recognition using cross-capture modality

Different algorithms with different capturing platforms are required to provide optimised solutions to different person recognition platforms. As an example, it is possible to capture individual in side-view in a surveillance environment where an adequate field of view is available [49, 50]. However, depth-based analysis is required to be perfromed on the frontal view in a narrow space. This research aims to provide independent cross-capture-modality gait recognition that is independent to the capturing platform for enrolment and testing, particularly for side-view and frontal-view depth image capturing systems, by finding answers to the following questions.

1. Are there any independent features that can be computed from frontal depth and side-view 2D?

2. How the back filling of the frontal contour in 2D affects the gait recognition performance?

3. Is it possible to achieve adequate gait recognition performances using the cross-capture modality features, compared to the capture-domain-specific features?

To answer these questions, independent features that can be generated by discarding the back leg information are explored. These new invariant features also will be investigated to provide view invariant solutions to 2D arbitrary view angle gait recognition.

Improvements to gait recognition using advanced feature optimisation techniques as a solution to unwanted appearance changes.

Based on the gait recognition framework shown in Figure 1.4, pre-processing before gait features are extracted and feature optimisation on computed features can also influence gait recognition performance, particularly in appearance-based techniques, as they are sensitive to appearance changes. This direction is explored by answering the following research questions,

1. Which parts of the silhouettes are mostly affected by external factors such as carrying bags, clothing etc. and which parts contribute to recognise an individual?

2. How do the dynamic and static nature of gait features influence gait recognition performance?

3. What is the impact of performing gait recognition on cleaned silhouettes using robust segmentation algorithms?

4. Can patch-based histogram methods be used to tolerate the segmentation noises and alignment issues in appearance-based methods? If so, how it can be optimised for a gait recognition context, with possible extension to 3D?

Part of the research aims to explore the recognition performance with advanced silhouette extraction techniques. It also proposes optimisation techniques that can handle the poor segmentation and image registration issues.

SRC based classifier in gait recognition context

Compressive sensing [51] and sparse representation [52] is becoming increasingly popular in many computer vision applications [53]. In particular, sparse representation-based classifiers are used as an effective classification method for face [52] and action recognition [54]. This research aims to investigate the applicability of SRC with an improved, locally discriminated input feature space in a gait recognition context, to improve the recognition performance by answering the following research questions,

1. Can a sparse representation-based classifier be used to improve the recognition performance in a gait recognition context?

2. How can the requirement of including all the variants in the dictionary to perform robust SRC be solved for the limited training data?

Algorithms developed throughout this dissertation are predominantly focused on appearance-based techniques due to their simplicity and excellent performance. However, initial investigation on model-based approaches are completed, and improvements over its baseline are achieved. Proposed approaches provide real world, gait recognition-based solutions to assist security applications in places including: where multiple-camera set-ups are available; portal-based biometric-gates fitted with frontal-depth sensors, cross-capture modality platforms; and unconstrained surveillance environments. CASIA [55], CMU MoBo [31], OULP [46] and an in-house developed depth gait database [56] are used to evaluate the proposed algorithms.

1.3 Original Contributions

Original contributions resulting from the research presented in this thesis include the following.

3D gait recognition using multi-view data

Current gait recognition approaches explore gait recognition on silhouettes captured in side-view, that is, when the person walks in a plane parallel to camera [48]. In 2D, side-view silhouettes represent the largest variation in gait dynamics as the legs extend to their maximum distances. When the view angle deviates, the gait representation on the silhouette gets reduced and recognition performance degrades heavily. This degradation is also contributed to, by misalignment of extracted features with the relevant parts.

To solve this issue, 3D representation of a subject is considered in this thesis. In large surveillance environments, it is possible to have multi-view images of a same person with the synchronised camera setup. Moving to 3D itself solves self-occlusion issues and loss of motion information in 2D.

Although computationally expensive, model-based techniques offer promise over appearance-based techniques, as they gather gait features directly and interpret gait dynamics in skeleton form. In this research, a fast 3D ellipsoidal-based gait recognition algorithm is proposed using the constructed 3D voxel model. A novel dynamic segmentation method is proposed to accurately segment right, left, upper and lower legs using Eigen-value decomposition technique. On each segmented region, ellipsoids are fitted, based on the energy distribution and ellipsoidal parameters that are extracted as static and dynamic gait features. Fourier components from computed features are used for similarity computation. The use of this 3D model approach is superior as it allows individual gait cycles of the left and right strides to be detected, segmented and modelled. The approach also significantly improves performance when there is a class mismatch between the gallery and probe sequences. This approach directly solves the limitations of view dependency and self-occlusion in existing ellipse fitting model-based approaches. An improvement of 15-20% true positive rate at false acceptance rate of 3% is achieved.

Even though the model-based approach in 3D improves the recognition rate than the similar model-based technique in 2D, the achieved recognition rate is comparatively lower than the 2D appearance-based technique, especially in side-view [57].

Therefore, appearance-based techniques in 3D are investigated and gait energy volume is proposed as a novel feature that accumulates 3D voxel volumes over a gait cycle and represents them in single gait energy volume. The proposed gait energy volumes have many advantages, simply by virtue of working in 3-dimensional space. This circumvents the issue of view dependency, as well as having no pose ambiguity (e.g. left- right limbs), no self-occlusion, and allows for easier segmentation of unwanted regions (e.g. hand movements). This appearance-based 3D feature achieved state-of-the-art recognition results compared to all other well-performing single-view and multi-view gait recognition techniques.

Frontal gait recognition using depth images

A frontal perspective has various advantages to that of the side, such as for use in narrow corridors, where the limited field-of-view of cameras may prevent the recording of complete gait cycles from the side. They can also be easily integrated into biometric portals. The addition of depth sensing ability to frontal capturing devices also enables more data to be captured than from the side, as there is no issue of self-occlusion. In fact, all gait-based information (kinematics) can be acquired from a frontal depth sequence. The proposed GEV-based appearance-based method is explored for applicability on frontal-depth. As only the front surface of the subject is visible, a partial voxel-model is created by taking the frontal surface reconstruction and filling to the back of the defined space along the depth axis. Using this, partial frontal-volumes frontal GEVs are computed as frontal gait features. To explore this, an in-house, frontal-depth database is developed, using Microsoft Kinect [58], which incorporates five different types of appearance changes and thus state-of-the-art recognition performance is achieved on this database.

In order to evaluate the effect of removing the back leg gait information in frontal GEV construction, synthesised frontal volumes are generated from full 3D voxel volumes and recognition performances of full 3D GEVs and frontal GEVs are compared. Based on the results, it has been shown that frontal depth has adequate gait information to distinguish an individual.

A cross-capture modality gait feature for frontal and side-view gait recognition

The main limitation in many gait recognition techniques is, that they can only work on a specific view-point and must utilise the same image capturing source. Overall, there are two extreme scenarios, frontal-view and side-view. Frontal-gait recognition is needed to perform human recognition in a concise space, and it can be better achieved by using the depth images. However, there are also situations where side view gait recognition is preferable, such as in a surveillance environment, where the distances involved may be unsuitable for many depth-sensing devices, or where depth-sensing hardware may simply not be present. These two different capture modalities operate in differing image domains and the gait features used in existing approaches are dependent in each domain. This prevents sharing of information without the use of view transformation models. In this research, a new gait energy-based feature, backfilled gait energy image (BGEI), that can be constructed from both side view silhouettes and frontal depth images, is proposed. This allows the feature to be applied across differing capturing systems using the same enrolled database, such as in a system using both frontal depth cameras mounted on biometric portals and general surveillance cameras.

The proposed BGEI feature is experimented for cross-capture modality (matching BGEI from side-view and frontal depth) on the DGD database and promising results are achieved. To evaluate the potentiality of BGEI compared to the relevant well-performing algorithms in each domain, BGEI is compared with GEI in side-view and with GEV in frontal depth. Even though drops in recognition results are noticed due to the nature of how the BGEI is built, performance degrades observed are in the acceptable margin.

This thesis also investigates the applicability of BGEI to perform view invariant gait recognition, as BGEI takes the frontal sketch of the silhouette. The main objective of this application is to provide arbitrary view angle gait recognition without any model training. To make a spatial match between inter-view BGEI, gait cycles for separated legs are computed and a spatial perspective transformation is applied to the silhouettes. With the promising initial results, future directions and guidance are provided for robust arbitrary view angle gait recognition.

Optimised pre-processing approaches for discriminative gait feature

As a part of the research, influence of silhouette extraction and pre-processing on gait recognition also has been explored particularly on appearance-based methods and came out with contributing findings and proposals. To compare the influence of segmentation, gait energy images are computed using cleaned silhouettes resulted from an optimised graph-cut based approach and the distributed silhouettes using simple background subtraction. Results shows that GEIs are sensitive to the segmentation noises as cleaned silhouettes achieve a better recognition rate. This is also a motivator to find a better optimiser for appearance-based methods, capable of withstanding segmentation noise.

Another issue in the appearance-based method is that it is sensitive to external appearance changes such as clothing, carrying bags etc. Since these changes can affect both static and dynamic regions of gait features, influence of the static and dynamic of gait towards recognition performance is explored. During the analyses, dynamic regions of the silhouettes provide consistence improved recognition rates with the existence of static-appearance changes and removing the static regions provides the solution to handle the static noises such as clothing. To provide a solution to both static and dynamic appearance changes (due to different types of clothing and carrying items), contribution of the GEI region with height from the bottom is explored, as the lower part of the body is less susceptible to external appearance changes. It has been found that significantly improved inter-class recognition results can be obtained when 40% of the silhouette height is used.

Sparse representation-based classifier for gait recognition

In most gait recognition techniques, the nearest neighbour classifier is used due to its simplicity. However representation of a test subject in terms of single training sample in this classifier is not feasible in an occluded/noisy environment. This issue is solved in a sparse representation-based classifier as it represents a test sample by considering all the possible contributions from within-class and between-class training data. In this thesis, SRC is explored in gait context and proposed as a well-performing classifier with significant improvements. SRC can work regardless of feature space as it tries to find the sparsest solution. However, it needs to have all the variants of a particular subject to be included in the training data. This is impractical with limited training data, particularly in a gait recognition context. To solve this issue, skewing the input feature space with class details using multiple discriminant analysis (MDA) is proposed. Improvement in recognition results are achieved with MDA-trained GEI features using SRC.

Local histogram feature descriptors

Though appearance based-methods have increased in popularity due to greater recognition performance and overall simplicity, they are still very sensitive to the segmentation noises and image alignment. Patch-based feature descriptors such as a histogram of oriented gradients (HOG) [59], and more commonly, local binary patterns (LBP [60]), are used with great effect to solve these issues, offering relatively high recognition performance at lower computational complexity.

Part of this research explores the effectiveness of HOG and an extended version of LBP, local directional pattern (LDP [61]), when applied to GEIs for gait recognition. A new feature descriptor, a histogram of weighted local directions (HWLD) is also proposed. All of the proposed feature descriptors manage to handle segmentation noises and show discriminating nature by achieving consistent recognition rate even without MDA transformation in SRC. The proposed histogram descriptor also enables possibility of extending it to 3D with affordable complexity.

1.4 Outline of Thesis

The remaining chapters of this thesis are structured as follows:

Chapter 2 provides a detailed survey of existing gait feature extraction techniques and evaluates them for their strengths and limitations. Techniques used for robust image segmentation, feature modelling and optimisation and classifiers that contribute to improve the gait recognition performance are also examined.

Chapter 3 describes the implementation of the gait recognition framework used in this thesis with a GEI-based baseline system. All required pre-processing, modelling and classification tasks are examined using popular methods chosen from the detailed literature review. It also explores the influence of silhouette extractions and dynamic/regional behaviour towards the recognition performance.

Chapter 4 illustrates 3D gait recognition in a multi-view environment. As a part of this, reconstruction of a 3D voxel model of a silhouette from synchronised multi-view data is explained. Detailed descriptions of the proposed 3D model-based, ellipsoidal feature descriptor and 3D extended, gait energy-based feature, gait energy volume are provided.

Chapter 5 explores the effect of discarding the back leg gait information in 3D to avoid the requirement of a multi-view set-up in GEV implementation. It also explains the frontal GEV implementation on depth images, captured from the Microsoft Kinect. The remainder of this chapter details the developed in-house frontal depth gait database and functionality of the developed frontal gait recognition GUI.

Chapter 6 outlines the implementation of SRC in a gait recognition context and the influence of a locally discriminated input feature space on recognition performance. It also illustrates feature optimisation in appearance-based feature extraction using patch-based descriptors and details the implementation and evaluation of the proposed HWLD algorithm.

Chapter 7 illustrates gait recognition in a cross-modality capture platform. Stability in recognition performance for back-filling the silhouettes in 2D and frontal 3D volumes are examined with the respective domain-specific feature, GEI and GEV. Based on this, the backfilled gait energy image is proposed as an independent capture-modality gait feature, particularly to accommodate frontal depth and side-view 2D.

Chapter 8 concludes this dissertation with a summary of the research as well as providing directions for future work.

Chapter 2

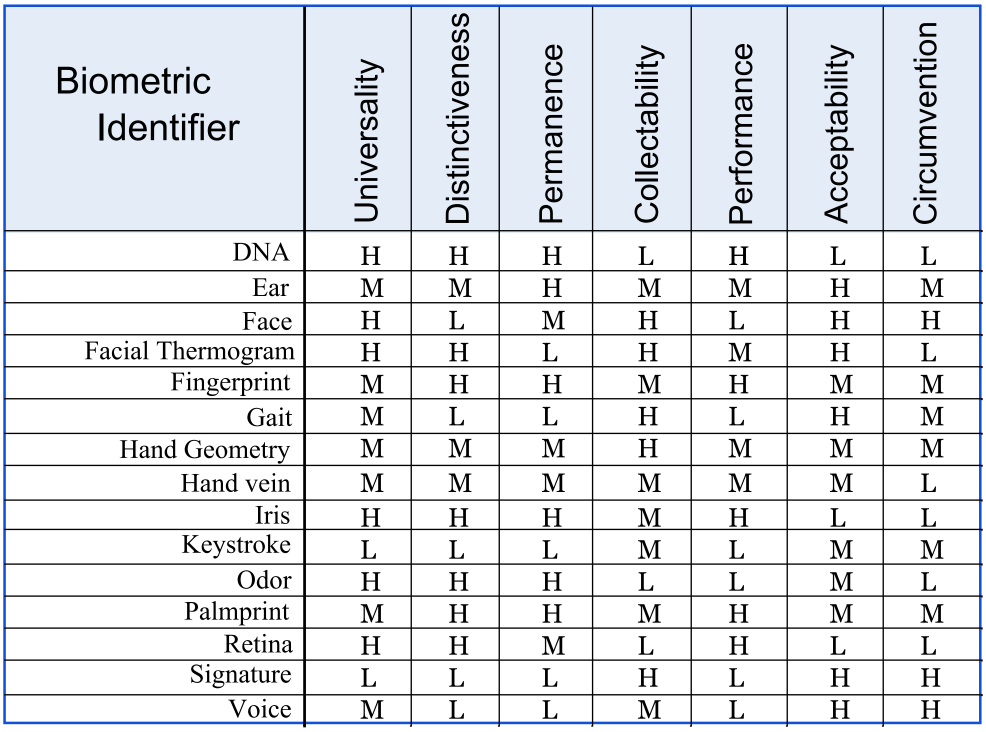
Literature Review

2.1 Introduction

There is considerable support for the notion that each person's gait is unique. Shakespeare mentioned in Twelfth Night [62] when Maria observes Malviolo: "By the colour of his beard, the shape of his leg, the manner of his gait, he shall find himself most pleasingly personated". The bio-mechanics literature makes similar observations: "A person will perform his walking pattern in a fairly repeatable way, sufficiently unique that it is possible to recognise a person at a distance by their gait" [63].

Table 2.1 illustrates the perception of the authors of [1] in considering "gait as a biometric" compared with other common biometrics based on several criteria such as universality, distinctiveness, acceptability *etc*. Gait has been shown as unique enough to identify an individual from several studies in literature, psychology and marker-based experiments results. However, this has not been evaluated with a large population of data and hence its' distinctiveness is rated as medium. Gait can be collected easily even without interacting with the subject and its acceptability index is high. Its universality index is medium as it cannot be collected when a person wears fully occluded clothes or is disabled. Since gait changes with age, the permanence characteristic is rated low.

Current recognition research achieves better performance in laboratory conditions and more focus on achieving the same in real world conditions. In this chapter, detailed descriptions of major contributed gait recognition research in the recent past is outlined.



Biometric Identifier	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial Thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand Geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

Table 2.1: Comparison of various biometric technologies based on the perception of the authors of [1]. High, Medium and Low are denoted by H, M and L.

2.1.1 Terminology

In this section, terminologies used in this thesis are explained for better understanding of the contents.

- **Gait Cycle /Gait Period**

  Gait cycle is description of the movements that take place during the walk from the time of one heel touching the ground till it re-touches itself on the ground, as illustrated in the Figure 2.1 [64]. There are phases in gait cycle known as stance and swing. Stance is the phase where the foot is touching the ground (classified as landing and taking off) and swing is where the limbs are moving through the air (classified as accelerated and decelerated).

- **Subjects/Sequences**

  The terms, ‘subjects’ or ‘sequences’ in this thesis, refer to the person IDs which are used for training and testing.

- **Gallery**

  Gallery is a collection of sequences of one or more gait cycles that are pre-

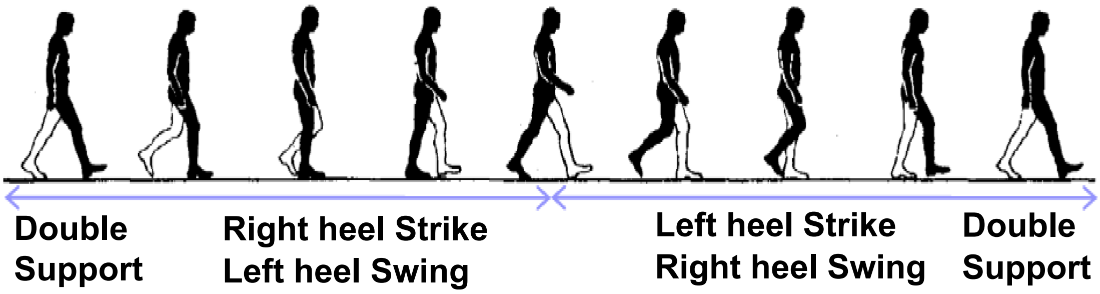


Figure 2.1: Different phases of gait cycle [64]. Frames between two similar stances with same pose represent a gait cycle.

enrolled in the database.

- **Probe**

  Probe sequences are a collection of one or more gait cycles of test subjects that need to be verified or identified.

- **Inter-class and intra-class**

  The terms, 'inter-class' and 'intra-class' describe the two different cases of experiments evaluated in this thesis. 'Inter-class' experiments are used to evaluate the algorithm between different gait challenging conditions of gallery and probe, while 'intra-class' is used to evaluate the gallery and probe in similar conditions.

2.2 Performance Analysis of Gait Recognition Methods

The performance of a system represents how accurately an algorithm classifies the subjects correctly. To estimate this, the samples from the database are divided into *gallery* and *probe* sets. The estimated rate of correct recognition is the fraction of the test set that the system classifies correctly. There are two common systems used in gait recognition - verification system and identification system.

1. Performance evaluation of an identification system.

   The cumulative match curve is used as a measure of $1 : m$ identification system performance. It judges the ranking capabilities of an identification system. Cumulative match score (CMS) is the output from the cumulative match characteristic curve and evaluates whether the correct answer is in the top k selected matches. This is evaluated in a closed-universe model (probe subject should be in the gallery). For the given probe sequence, the gallery sequences are ranked according to their similarity (distance to the probe). Assume P numbers of probes are selected for testing, and C_k numbers of probes are correctly classified within the top matches, then the cumulative match score is computed as,

$$\text{CMS}(k) = \frac{C_k}{P}. \tag{2.1}$$

Generally, k is chosen to be in the range 1 to 3. It shows the probability that a given user appears in different sized candidate lists. The faster the CMC curve approaches 1, indicating that the user always appears in the candidate list of specified size, the better the matching algorithm. Figure 2.2(b) shows typical cmc curves that show the changes of CMS with accumulating ranks.

2. Performance evaluation of a verification system.
The performance of any 1:1 biometric verification systems can be evaluated by expressing the trade-off between false acceptance rate (FAR) and false rejection rate (FRR). FAR is the percentage of accepted non-genuine individuals with the total acceptance made by the system, defined as,

$$\text{FAR} = \frac{\text{FN}}{\text{FN} + \text{TP}}, \tag{2.2}$$

where FN is false negative (non-genuine individual accepted by the system) and TP is true positive (genuine individual accepted by the system). FRR is a percentage of rejected genuine individuals compared to total rejects made by the system, defined as,

$$\text{FRR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \tag{2.3}$$

where FP is false positive (genuine person rejected by system) and TN is true negative (non-genuine individual rejected by system). Ideal human identification system requires the recognition performance with, both FAR and FRR at zero level. However, achieving this is impractical in real world application scenario. In this case, type of the application determines the required level of FAR and FRR. Application, such as providing access control, always prefers to keep FAR lower as possible, because level of non-genuine person's access needs to be minimised. Applications in the forensic analyses need lower false reject rate, as it can't miss any of the true individual connected to a crime activity.

Receiver operating characteristic (ROC) curves are a well-accepted measure to express the performance of 1:1 matches. An ROC curve plots, parametrically as

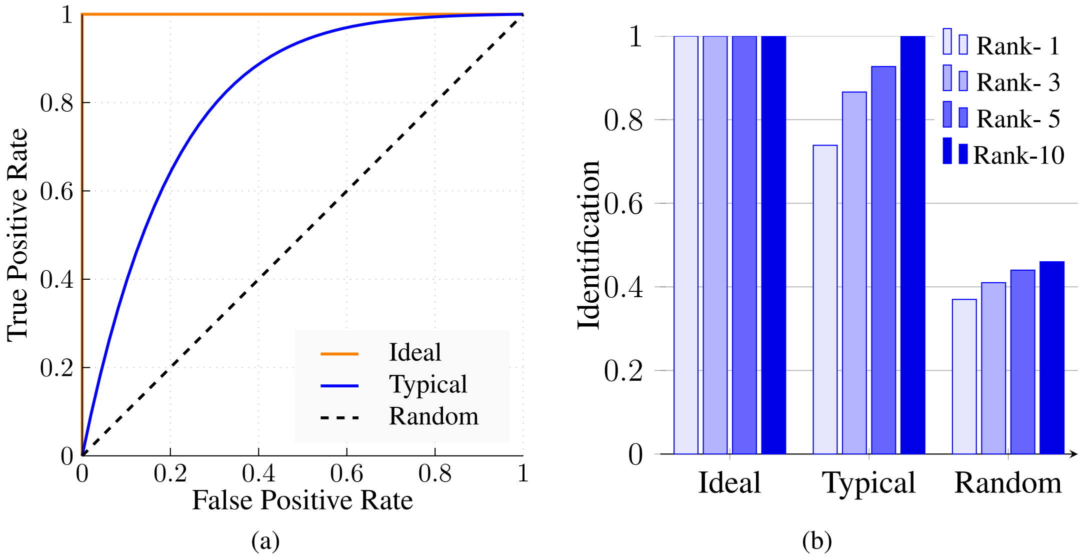


Figure 2.2: Typical performance measuring curves used in gait recognition research. (a) Receiver operating characteristic (ROC) curves and (b) cumulative match scores (CMS) for competitive ranks.

a function of the decision threshold, FAR (impostor attempts accepted) on the x-axis, against the corresponding true acceptance rate (TAR - genuine attempts accepted) on the y-axis, where $\text{TAR} = (1 - \text{FRR})$. In general, if the number of true positives (sensitivity) increases, the number of false negatives (specificity) also increases. This relationship between sensitivity and specificity is described in the ROC curve as shown in Figure 2.2(a). ROC curves are threshold independent, allowing performance comparison of different systems under similar conditions, or of a single system under differing conditions.

2.3 Available Databases

From the early nineties, researchers have been developing algorithms to make use of gait for human identification. However, early research has suffered from a lack of databases to benchmark and improve methods in realistic conditions. During 2001-2004, DARPA [65] has distributed real world gait data collected at NIST as a part of a *human ID at distance (humanID)* challenge program and each contributing team from MIT, Maryland, Southampton, CMU and USF were tasked to analyse the data. As an initial step, a SOTON database [66] was developed in 2001 with 33 subjects with five covariates. This was extended by Sarkar*et al.* [67] with the number of subjects

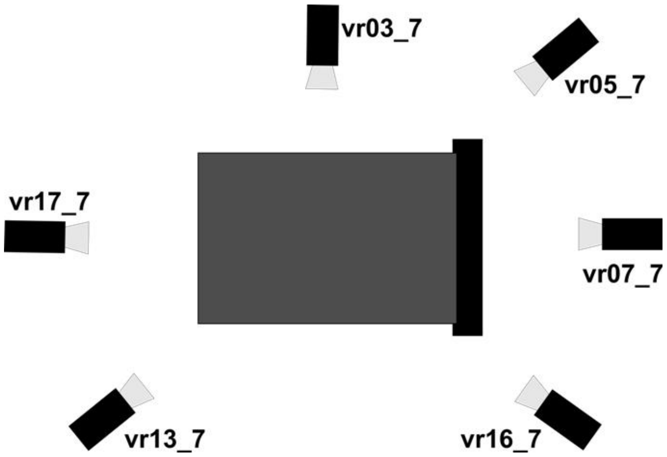


Figure 2.3: Camera set-up of the CMU MoBo multi-view dataset. Six cameras are positioned to cover the complete field-of-view of the walking person on the treadmill.

increased to 121 and changes in time also added as another factor. Following this, a number of databases have been developed with an increasing rate of research in gait recognition. The remainder of this section describes the popular databases that are currently available and used for gait recognition research.

1. CMU MoBo Database [31]
   The database has 25 subjects with six views and four walking styles. A large amount of image data is available to train and test algorithms (8000 images per subject). The main drawback of this database is that all the data is from an indoor environment (collected from a treadmill). Multiple cameras were positioned to capture different views, as illustrated in Figure 2.3.

2. University of Southampton- Human ID at a Distance (SOTON) Database [66]
   This is a database specifically developed to address the current limitations in gait recognition algorithms. It contains two types of datasets – a large dataset with more than 100 subjects and a small dataset with only 10 subjects. The large dataset has two viewpoints (frontal and oblique) and contains subjects in both outdoor and indoor environments and on a treadmill. The small dataset is extensive with respect to conditions such as the type of footwear, clothes and surface.

3. UMD (University of Maryland) Database [68]
   This is a challenging database, which contains outdoor surveillance data. It has gait sequences for individuals with diversity in terms of gender, age, ethnicity,

etc. It includes sequences with different clothing on different days. The data was captured in a realistic scenario with uncontrolled illumination. The database contains walking sequences of four different poses. The database can be divided into three individual datasets.

The first dataset involves 25 subjects captured in four views. The second dataset consists of videos of 55 participants walking along a T-shape pathway. Two cameras with their optical axes orthogonal to each other captured the gait of each participant. The third dataset contains only 12 participants and has people walking at angles of 0, 15, 30, 45, and 60 degrees to the camera.

4. CMU (Carnegie Mellon University) Graphics Lab Motion Database [69]
   This database is freely available and contains 12605 trials in six categories. Each of these categories is subdivided into 23 groups. All of them are in the frame rate of 120.
   Example categorisation of the database includes: Human Interaction (two subjects), Interaction with environment (playground, uneven terrain etc.), locomotion (running, walking), physical activities and sports (basketball, dance), situations and scenarios (common behaviours and expressions, etc.) and test motions.

5. Georgia Tech Database [70]
   This database contains 3D motion capture trajectories with 20 participants. For some data, different view angles and outdoor data are available. The main drawback here is the database is relatively small and it is difficult to test the algorithm at different levels of difficulty.

6. MIT (Massachusetts Institute of Technology) Database [71]
   This database has 194 sequences and allows testing of an algorithm's robustness to clothing changes (multiple days). This data provides a single point of view (fronto-parallel).

7. CASIA Gait Database [55]
   This is a newly developed challenge database for the current approaches in gait recognition. It is available in three datasets. Dataset A consists of 20 subjects. Each subject has 12 image sequences (length of 37 to 127 frames), four sequences for each of the three directions (parallel, 45 degrees and 90 degrees to the image plane). Dataset B (large multi view gait database) has 124 subjects for 11 view angles (from 0° to 180° stepped by 18°) and for each view angles

with four sets of normal condition and two sets of subjects with bag and subjects wearing coat. Dataset C was collected by an infra-red (thermal) camera and contains 153 subjects and four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag.

8. OULP Gait Database [46]
   OULP Gait database is a recently developed database with more than 3000 subjects. Subjects are included from multiple ages and genders and captured using two cameras to accommodate two sequences for each subject in each view for the angles of $55°$ to $85°$.

2.3.1 Summary

Table 2.2 compares the available gait databases and the included variations with their meta data. In our research, three popular databases, CASIA database, OULP database, CMU MoBo database, are obtained to benchmark our algorithms in three directions. Since the CASIA database has multiple gait challenging conditions on a large number of subjects, it is used to analyse the effect of the appearance changes and variations on gait challenging conditions. To explore the research in 3D direction, CMU MoBo is used as it is the only available synchronised multi-view database. For robust experiments on large scale data, the OULP database is used.

2.4 Current State of Art

Current approaches in gait recognition are broadly categorised into appearance-based and model-based. Appearance-based approaches are based on the moving shape of the silhouette or integration of information from the moving shape and the motion. In appearance-based approaches, gait information is extracted by either structural model or motion model.

Figure 2.4 depicts the meta-analysis (combination of different information of affecting factors) of gait recognition with average results for state-of-the-art algorithms on the most common databases, such as CMU MoBo, SOTON and CASIA. It can be seen that an indoor environment achieves an identification rate of 80%, however the identification rate drops over 10% in outdoor environments. The most significant drops occur due to changes in the surface and for large test session intervals (6 months).

Database	No of Subs.	No of seqs.	Environment	Variations
SOTON small	12	-	indoor	carrying, clothing, shoe, view
SOTON large	115	2128	indoor, outdoor, treadmill	view
USF	122	1870	outdoor	views, surface, shoe, carrying, time
UMD -1	25	100	outdoor	4 views
UMD -2	55	220	outdoor	2 views
MIT	24	225	indoor	side-view
CMU MoBo	25	600	indoor, treadmill	6 viewpoints, speed, incline, carrying
CASIA dataset B	124	13640	indoor	11 viewpoints, carrying, clothing
OULP	3000-4000	-	indoor	views

Table 2.2: Summary of used gait databases. Size of the database in terms of number of subjects, included variations and captured conditions are compared.

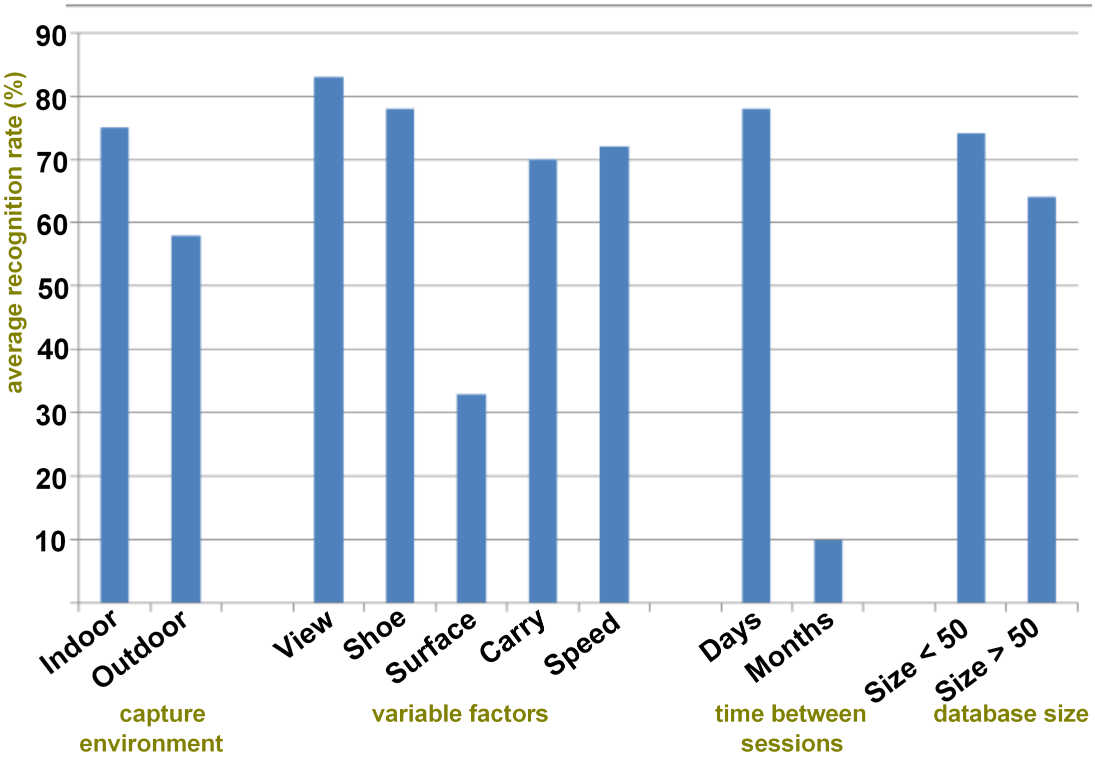


Figure 2.4: Meta-analysis of gait identification rates for different conditions [29, 72]

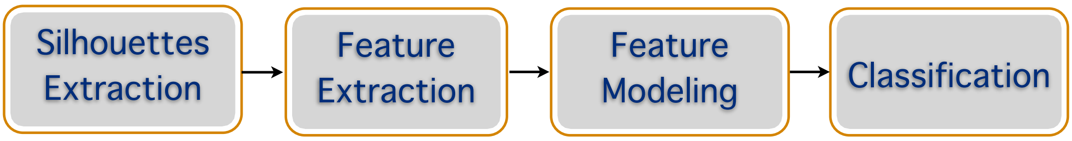


Figure 2.5: Major components of gait recognition algorithms. At first, silhouettes are extracted and pre-processed using background subtraction algorithms. Then, gait features are computed from the segmented silhouettes and these gait features are modelled to encode better distinguishable elements. Finally, these modelled-features of the test subject are compared on the pre-enrolled database to claim the identity.

2.5 Gait Recognition Framework

Figure 2.5 describes a typical framework for gait recognition. Initially, segmentation processes are applied to extract the human silhouettes from the video sequences (see Section 2.6). A number of post-processing tasks including normalisation of the silhouettes are applied to the extracted silhouettes to register them on a common platform. Then gait features are extracted and modelled to encode the individual-specific gait information. Finally, classification is performed to compute the similarity distance that will be used to claim individual's identity. Within all of these processes in the framework, the most important step is extracting appropriate gait feature that can uniquely distinguish subjects. Existing feature extraction methods can generally be classified as appearance-based methods (see Section 2.7.1) and model-based methods (see Section 2.7.2). Both appearance-based or model-based methods can be applied to 2D or 3D data sources. This section discusses the research pertaining to the major components in the gait recognition framework: silhouette segmentation, feature extraction, feature modelling and classification.

2.6 Silhouette Extraction

As in many computer vision applications, identifying the human silhouette is a fundamental task in gait recognition [73]. A common approach is to perform background subtraction, which identifies the human silhouette as a moving object within the scene. This is very effective, where few objects with significant size move. The detection becomes more difficult where additional subjects, such as cars, move together with humans in the scene. In this case, a simple size-threshold can be applied to the motion mask to distinguish humans from other moving objects. There are many challenges in

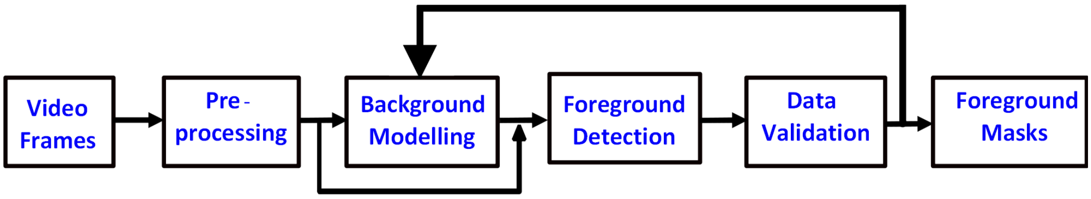


Figure 2.6: Flow diagram of background subtraction algorithms. Silhouettes are pre-processed for initial registration and background is modelled. Foregrounds are predicted using the trained model. This process is iteratively performed on training data to fine-tune the background model to generate a robust foreground mask.

developing a good background subtraction algorithm. It must be robust against changes in illumination and needs to avoid detecting non-stationary background objects such as clouds, vehicles, moving leaves and shadows [50, 74]. Typical approach of analysing these movements in the video is to use optical flow [75, 76].

A typical background subtraction algorithm follows a process illustrated by the flow diagram in Figure 2.6. This consists of four major steps known as pre-processing, background modelling, foreground detection and data validation. A number of background subtraction techniques have been developed over the past. Several popular techniques are outlined in this section.

2.6.1 Traditional Approaches

1. **Frame Differencing [73]**

   Frame differencing is a simple motion segmentation algorithm. It operates by computing the difference between consecutive frames and applying a threshold,

   $$|frame_i - frame_{i-1}| < T_h. \tag{2.4}$$

   Here, estimated background is just the previous frame. This method is only suitable for a particular frame rate and object speed.

2. **Background as Average or the Median**

   Cucchiara *et al.* [77] and Velastin *et al.* [78] proposed a method to estimate the background as the average or median of the previous n frames. This is a more robust method than frame differencing. Even though this method is fast, it consumes a large amount of memory. A modification made to reduce the memory requirements is to compute the background as a running average. Each pixel location in the background is modelled based on the recent history of that pixel (by means of the average, weighted average or chronological average).

3. **Running Gaussian Average**
The approach of applying a Gaussian distribution to the background is a significant turning point in the background subtraction techniques. Wren *et al.* [79] used a single Gaussian to represent the background. A Gaussian distribution is fitted to the image histogram to form the background probability density function (PDF). Then the background PDF is updated using the running average method. Let the Gaussian distribution have mean μ and the variance σ^2, then the running average method will update the background PDF continuously as,

$$\begin{aligned} \mu_{t+1} &= \alpha * F_t + (1-\alpha) * \mu_t, \\ \sigma^2_{t+1} &= \alpha(F_t - \mu_t)^2 + (1-\alpha)\sigma^2_t, \\ |F - \mu| &> T_h. \end{aligned} \tag{2.5}$$

T_h can be chosen as $k\sigma$, where F is the image pixel, α is learning rate and t represents timing sequence.

2.6.2 Methods Based on Dynamic Models

1. **Mixture of Gaussian and its extensions**
Mixture of Gaussian
A single Gaussian covers the changes in the background geometry. However, the changes in the background may appear at a rate faster than that of the background update. Outdoor scenes with moving trees, rain or changes in clouds are examples. In these cases, a single valued background model is not adequate. The concept of representing each pixel in the image by a Mixture of Gaussian was introduced by Stauffer *et al.* [80].

They express the probability of observing a certain pixel value x at time t using a mixture of Gaussian weighted with w_i as,

$$P(x_t) = \sum_{i=1}^{k} w_{i,t}\eta(x_t - \mu_{i,t}, \Sigma_{i,t}). \tag{2.6}$$

Each pixel is described using k Gaussian distributions. Commonly k is set to be between 3 to 5. At each new frame, some of the Gaussian match the current value (within a variation of less than 2.5σ). μ_i, σ_i are updated using running average. The criterion for separating background and foreground is defined by ranking the distribution based on the ratio between their amplitude, w_i, and standard deviation, σ^2. The first set (let's say B) of distributions in the ranking order satisfying,

$$\sum_{i}^{B} w_i > T_h, \tag{2.7}$$

are realised as background. Here, T_h is a manually defined threshold.

Kernel density estimation (KDE)
An improved version of the mixture of Gaussian, Kernel Density Estimation (KDE), is proposed by Elgammal *et al.* [81]. This is a kernel-based approach where the background pixel is represented by the individual pixels of the last N frames. The background PDF $P(x_t)$ of pixel x at time t is given as a sum of Gaussian kernels centred on the most recent N background values. If pixel x_t satisfies,

$$P(x_t) < T_h, \tag{2.8}$$

then it is classified as a foreground pixel, where T_h is threshold. This approach requires more memory and time to find the kernel values.

2. **Sequential Kernel Density Approximation (SKDA)**
Han *et al.* [82] used mean shift vector to estimate the modes from samples. Mean shift vector is an effective gradient-ascent technique able to detect the main modes of the true PDF directly from the sample data with a minimum set of assumptions. However, it has a very high computational cost, since it is an iterative technique and it requires convergence over the whole data space. As such, it is not immediately applicable to modelling background PDFs at the pixel level. Therefore, Han used this at the initialisation time only. After that, modes are propagated by adapting them with the new samples as,

$$PDF(x) = \alpha(\text{new model}) + (1-\alpha)\Sigma(\text{existing models}). \tag{2.9}$$

The main advantage of SKDA is that the number of modes is not defined a *priori* and that the combined method works faster and consumes less memory.

3. **Motion segmentation with Graph Cuts**
   Chen *et al.* proposed a method for combining motion segmentation with graph cuts to improve the segmentation performance of moving objects in video sequences in [83–86]. Here, the gradients have been used in conjunction with intensity and colour, to provide a robust segmentation of motion. Then, graph cuts (see Section 3.4.1 for more details) are applied to the output of the motion image to remove spurious motion and fill-in motion that could not be detected. This technique was again improved in [84] by simultaneously computing two background models as input into the graph cut algorithm . Improved results are obtained by extending the base motion segmentation system to extract both a foreground and background model. The segmentation results using these graph cuts work well, since this can handle the regions of uncertainty better when foreground pixels appear similar to those of the background, or when the background changes significantly due to shadows and reflections. Therefore, throughout the research, this method is used for silhouette extractions for the gait analysis.

2.7 Feature Extraction

Two approaches are commonly used to extract features for gait recognition, namely model-based analysis and appearance-based analysis (also known as model-free or silhouette-based analysis). Earlier approaches are mainly appearance-based methods. However, these methods are severely affected by the challenges in outdoor environments. Model-based approaches give better results on data even with higher noise, due to illumination and surface effects. Early versions of model-based approaches were developed based on static parameters such as stride length, length of limbs etc. Later, research started to work on the skeletonisation of silhouettes and used dynamic features such as joint angle trajectories. Recently, the research focus has shifted to multi-view 3D analysis to model silhouettes from any view point.

2.7.1 Appearance Based Approaches

There are a number of techniques that have been developed for gait feature extraction based on the silhouette appearance. The remainder of this section discusses several popular techniques.

Spatio-temporal silhouette analysis

In 2001, Phillips *et al.* [87] developed an algorithm based on the entire silhouette in the fronto-parallel view. This approach is parameter-based and requires manually input threshold values. Silhouettes are directly compared between probe and gallery images. Probe sequences are subdivided into a small number of subsets (manually defined gait period) to compare with gallery images. Both CMC and ROC curves were used to evaluate the algorithm on SOTON [66]. When the changes are allowed only in view, a CCR of 86% was achieved. They achieved a poor result of 42% when there is a change in surface. In 2005, Sakar *et al.* [67] extended the work of Phillips *et al.* [87] by a parameter free model by extracting the silhouette by means of iterative expectation maximization. The CMU MoBo database [31] was used for their experiments. They normalised the similarity values based on the Median of the Absolute deviation (MAD) and achieved a better CCR of 94% with only changes in view and 80% with changes in surface. However when all factors (clothes, shoe types, surface, time) were allowed to change, they could only achieve a CCR of 27%.

Collins *et al.* [88] used the shape of the human to identify key frames. For side views of a walking person, the width of the silhouette was plotted as function of time. Key frames were selected at the peaks and valleys, which occurred periodically on the plotted function. Consecutive pairs of peaks and valleys were used to represent a single stride. These key frames roughly corresponded to the following gait labels: right double support (both legs spread and touching the ground, right leg is in front), right mid-stance (legs are closest together with the swinging left leg just passing the planted right foot), left double support, and left mid-stance. For the front view of walking, the same procedure was done, except height was plotted as a function of time instead of width. This technique reduces the computation time and cost compared to [87]. They have tested the algorithm in three conditions in four different databases. A recognition rate of more than 90% was achieved for intra-class tests on most common databases. Inter-class tests on CMU MoBo also achieved more than 90%.

Extending shape descriptions to moving objects

1. **Area mask**
   The area masks derive a measure of the change in area of a selected region of a silhouette. Foster *et al.* [89] uses masking functions to measure area as a time varying signal from a sequence of silhouettes of a walking subject. This tech-

nique is an extension of [67], that combines the simplicity of [67] area measure with the specificity of the selected (masked) area. Different types of masks are used to gather gait signal from each parts (limbs, torso, etc.) of human silhouette separately. Combination of the gait signal using these different masks achieved the recognition rate of 76% on extended SOTON [66] database with 114 subjects.

2. **Gait symmetry**
   The discrete symmetry operators are one of the most widely-deployed symmetry operators in computer vision techniques [90]. They use edge maps of images from the sequences of subject silhouettes to assign symmetry magnitude and orientation to image points, accumulated at the midpoint of each analysed pair of points. Nixon *et al.* [90] use this symmetry operator to extract the gait feature from the extracted silhouette. Initially, the Sobel operator is applied to the extracted silhouette to derive the edge-map. A threshold is applied to the edge-map, so as to set all points beneath a chosen threshold to zero, to reduce noise or remove edges with weak strength. This reduces the amount of computation in the symmetry calculation. The symmetry operator is then applied to give the symmetry map as shown in Figure 2.7. The Gait signature (GS) of the sequence of images is obtained by averaging all the symmetry maps as shown below,

   $$GS = \frac{1}{N}\sum_{i=1}^{N} S_j, \tag{2.10}$$

   where S_j is the j^{th} symmetry map in the sequence and N is the number of symmetry maps for a particular sequence.

3. Velocity moments
   Shutler *et al.* [91] developed Zernike velocity moments to describe an object, not only by its shape, but also by its motion throughout an image sequence. Here, images are stacked together into a three dimensional XYT (space plus time) block, and then a 3D descriptor is applied to this data. By treating time as a depth axis, velocity information is represented using conventional 3D moments. Velocity moments are based around the Centre Of Mass (COM) description and are primarily designed to describe a moving and/or changing shape in an image sequence. This helps to describe the gait in terms of the structure of the moving shape together with motion information. The generalised velocity moment from

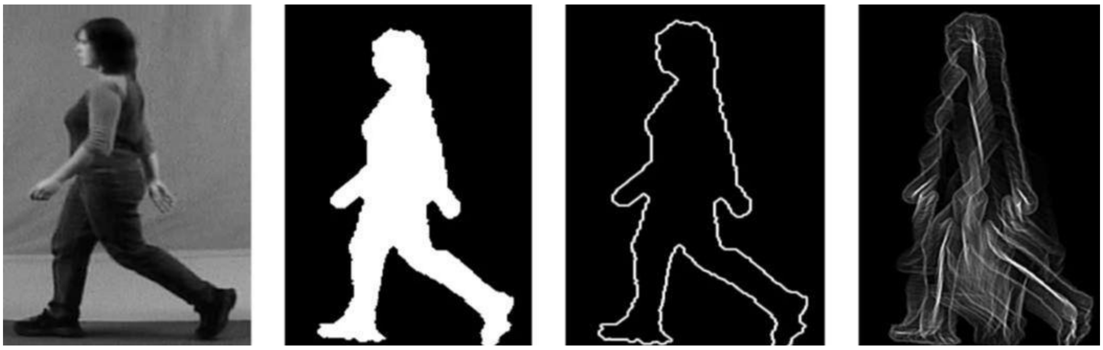


Figure 2.7: Symmetry map computation by Nixon *et al.* [90]. From left to right original image, extracted silhouette, edge-map after Sobel operator is applied and symmetry map are shown.

a sequence of images is,

$$VM_{m,n,\alpha,\gamma} = \sum_{i=2}^{n} \sum x \sum yU(i,\alpha,\gamma)S(i,m,n)P_{i_{x,y}}, \tag{2.11}$$

for image i and the moments are of order m, n, α, γ. $S(i, m, n)$ is the standard spatial moment and the velocity is introduced through $U(i, \alpha, \gamma)$ which are calculated from the differences between consecutive COMs in the image sequence. The experiment is done only for the intra-class case on their own database with 42 sequences including six subjects; a recognition rate of 100% was achieved in this test case.

Procrustes shape analysis (PSA)

Wang *et al.* [92] used Procrustes Shape Analysis to extract the gait signature in the form of the mean shape from the complex configuration representation of silhouette images. Initially, the boundary of the silhouette images are obtained using a border following algorithm based on connectivity and the shape centroid is calculated by computing the mean (x_c, y_c) of the boundary pixel coordinates. Then, boundary pixel points are unwrapped, based on the centre coordinates and each shape is described as a vector of ordered complex number (termed the configuration of the shape) with N_b number of boundary points as follows,

$$u = \left[z_1, z_2, ..., z_i, ..., z_{N_b}\right]^T, \tag{2.12}$$

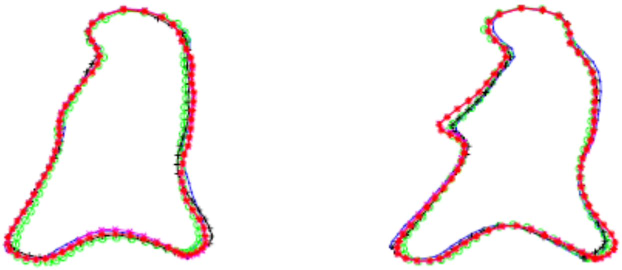


Figure 2.8: Procrustes mean shapes of two different persons. Number of sequences are shown in different colours.

where, $z_i = (x_i - x_c) + j \times (y_i - y_c)$. The full Procrustes distance $d_F(u_1, u_2)$ between two centralised configurations (u_1, u_2) can be defined as in Equation 2.13. This Procrustes distance allows the comparison of two shapes independent of position, scale and rotation,

$$d_F(u_1, u_2) = 1 - \frac{|u_1^* u_2|^2}{\|u_1\|^2 \|u_2\|^2}. \tag{2.13}$$

Superscript $*$ represents the complex conjugation transpose and $0 < d_F < 1$. Based on this, the mean shape S_u of the given n shapes can be computed as follows,

$$S_u = \Sigma_{i=1}^{n} (u_i u_i^*)/(u_i^* u_i). \tag{2.14}$$

The Procrustes mean shape ($\hat{u}$) is constructed using the eigenvectors corresponding to greatest eigenvalues of S_u. Figure 2.8 shows the Procrustes mean shape of two different subjects with different sequences. The similarity between two gait sequences is computed using Procrustes mean shape distance (MSD). Classification is done using kNN (see Section 2.9) at $k = 1$, $k = 3$ on the CASIA [55] (dataset: A). The kNN gave a better average recognition rate of 88% for different view angles. This approach was extended by Zhang *et al.* [93] using Shape Context (SC) instead of MSD. A set of vectors as shown in Equation 2.12 with a defined number of points (N), originating from a random point p_i to all other points on the shape, expresses the configuration of the entire shape relative to the reference point. Shape Context of point p_i is defined as a coarse histogram (h_i) relative to the coordinates of the remaining $N - 1$ points and can be calculated as follows,

$$h_i(k) = \text{ Number of } \left\{ q \neq pi : (q - pi) \in bin(k) \right\}. \tag{2.15}$$

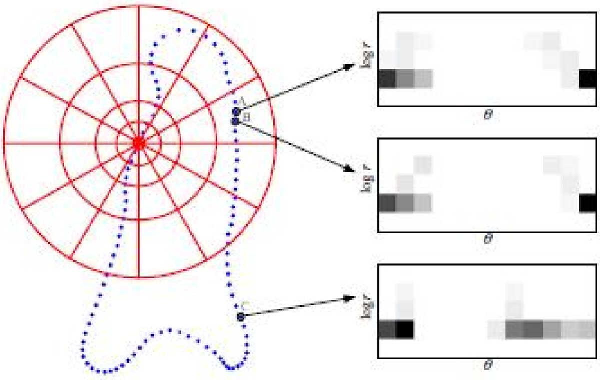


Figure 2.9: Computation of shape context (SC) for procrustes mean shape [93]. The circles and lines attached on the PMS are the diagram of log-polar histogram bins used in computing the SCs. Here, 5 bins are used for log-radius and 12 bins are used for θ. Examples of SCs for three points A, B, and C are shown in the right column (Dark = large value).

Equation 2.15 counts the number of boundary points within each sector or bin to form the SC. The method of computing SC is illustrated in Figure 2.9.

FED (Frame exemplar distance) with HMM

Kale *et al.* [94] introduced a Hidden Markov Model (HMM) based approach. The concept of FED (Frame to Exemplar Distance) reduces the dimensionality of the feature vector as an indirect approach. They have defined four exemplars as the key frames that represent a gait period, as in [88]. Then if $x(t)$ is the feature vector extracted from the image at time t, the distance from $x(t)$ to the corresponding exemplars is computed as the FED vector. They chose two types of feature vectors: width of the extracted silhouettes and the entire silhouette itself.

Choice of these feature vectors depends on the available image quality. When higher resolution images are available, the width of the silhouettes are used and the entire silhouette is used for noise-corrupted images. Again, there are two proposed feature comparisons: direct approach and indirect approach. In the direct approach, the feature vectors are used as is, without any synthesizing. Alternatively, in the indirect approach, FED (Frame to Exemplar Distance) is computed for comparison.

The Hidden Markov Model was used as a classifier (will be described in Section 2.9.3) for both techniques and showed that the proposed FED-based method out performs the direct approach. The algorithm was tested in CMU MoBo [31], UMD [68] and SOTON [66]. With the 122 people SOTON database [66], a CMS of 89% was achieved with only changes in view. However, a CMS of only 35% was achieved when changes in surfaces were allowed. With all the co-variates (type of shoes, surface, time, view) a CMS of 15% was achieved.

Self similarity plot (SSP)

Benabdelkader *et al.* [95] proposed the Self Similarity Plot (SSP) to encode the projection of gait dynamics. Initially, consecutive image frames (N) of a particular person are scaled to a pre-determined height. Image self-similarity of the particular person is computed as follows,

$$S(t_1, t_2) = \sum_{(x,y)\in B_{t_1}} |O_{t_1}(x, y) - O_{t_2}(x, y)|, \tag{2.16}$$

where $1 \leq t_1, t_2 \leq N$, B_{t_1} is the bounding box of the person in frame t_1, and O_{t_1}, O_{t_2},-.., O_{t_N} are scaled image templates of the person. S is again translated over small radius r to address the tracking errors and minimal S' was computed as,

$$S'(t_1, t_2) = \min_{dx,dy<r} \sum_{(x,y)\in B_{t_1}} |O_{t_1}(x + dx, y + dy) - O_{t_2}(x, y)|. \tag{2.17}$$

The above computed self-similarity values are linearly scaled for visualisation in the grey scale intensity range $[0, 255]$ and plotted. These self-similarity plots (SSP) were chosen as the feature vector and an eigen face approach [96] was used as the gait classifier. PCA, LDA and Subspace LDA were used to reduce the dimensionality of the feature vector. This technique is more robust to non-white noise and nonlinear amplitude modulations compared to methods based on Fourier analysis 2.8.4. They tested their algorithm in Little and Boyd [97] (40 image sequences with six different subjects). They used the kNN classifier for recognition and achieved a CMS of 93% at rank 1.

Gait energy image

The Gait Energy Image (GEI) [98] is the average silhouette taken over a single gait period, enabling the temporal information of gait to be encoded in a single frame. As such, it is less sensitive to silhouette noise in individual frames and is less affected by varying gait periods. Given the pre-processed binary gait silhouette images $I_t(x, y)$ at time t in a sequence, the gray-level Gait Energy Image is defined as follows,

$$G(x, y) = \frac{1}{n} \sum_{t=1}^{n} I_t(x, y), \tag{2.18}$$

where n is the number of frames in the complete gait cycle(s) of a silhouette sequence, t is the frame number within the gait cycle and x and y are $2D$ image coordinates.

Since each silhouette image represents the space normalised energy image of a human walking at a particular time, the GEI is the time normalised accumulation of these energy images of the human walking for a complete gait cycle. In GEI, high intensity pixels correspond to a greater frequency than that at which the silhouette is observed at that position.

Xiang *et al.* [99] used MGEI (Mean Gait Energy Image) instead of GEI and used Kernel Principal Analysis to define a low dimensional gait feature. Consider the GEIs, $G_k(x, y)$, where $k = 1, 2...n$, (x, y) are the $2D$ image coordinates and n is the number of complete gait cycles in the sequence, then MGEI can be computed as in the Equation 2.19 to represent the major shapes of silhouettes and mean gait characteristics of their change over gait cycles,

$$MGEI(x, y) = \frac{1}{n} \sum_{k=1}^{n} G_k(x, y). \tag{2.19}$$

This represents a space-normalised energy image of a subject walking at a given instance and the probability accumulative energy image of a subject walking over a complete cycle. The Euclidean distance of the weighted reciprocal was taken as the classifier. Three different experiments were performed under different levels of covariant affects, such as changes in carrying goods and clothes. The first and second experiments used the CASIA gait database [55], while the third experiment used the SOTON database [66] and the level of changes in the covariates increased from the first experiment to the third. The first, second and third experiments yielded a CMS of 92%, 52% and 57% respectively at rank 5.

Modified and improved versions of GEI in the form of energy deviation image (EDI) [100], and enhance gait energy image (EGEI) [101] are now used with different types of classification techniques and have shown promising improvements in the results.

2.7.2 Model Based Approaches

Since medical studies showed that joint angle trajectories are unique to each person [102], researchers have begun to use extracted joint angle trajectories for gait recognition approaches. They found that model-based approaches are suitable for this extraction process. Model-based methods are more robust to a variety of factors (changes in appearance of walking person due to clothing, carrying goods, segmentation noises etc.) and typically yield better recognition results compared to appearance-based approaches in inter-class conditions. However, these model-based methods require accurate model fitting and registration for robust gait representation and that is highly computationally expensive and possible when clean data is available. In a typical model-based approach, a structural model and a motion model are required to serve as the basis for tracking and feature (moving human) extraction [57]. Popular existing techniques based on the model-based approach are analysed in the remainder of this section.

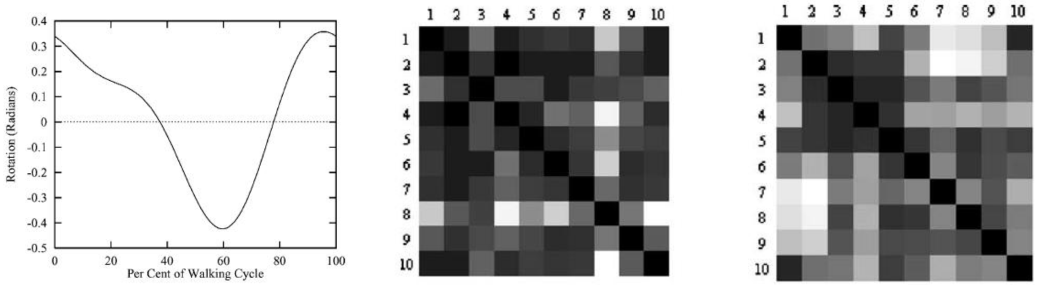
Structural models

The structural model represents the human body parts – the head, torso, hip, thigh, knee and ankle – with primitive shapes (cylinders, cones, and blobs) by measurements of length, width and position. Bobik *et al.* [101] showed that significant performance improvements can be achieved using stride as a gait feature. They used the action of walking to derive relative body parameters, which described the subject's body and stride. They analysed within-class and between-class variation to determine the potency of the technique on motion captured data. They found that the relative body parameters appeared to have greater discriminatory power than the stride parameters.

Fusion of structural and motion Models

A motion model describes the kinematics or the dynamics of the motion of each body part [57]. Kinematics generally describe how the subject changes position with time, without considering the effect of masses and forces, whereas dynamics take into account the forces that act upon these body masses and hence the resultant motion. When developing a motion model, the constraints of gait, such as the dependency of neighbouring joints and the limit of motion in terms of range and direction, has to be understood.

Most recently, it was found that fusion methods of combining motion parameters



(a) (b) (c)

Figure 2.10: Algorithm proposed by Cunado *et al.* [103] (a): Hip rotation pattern extracted with the thigh model. (b): Similarity matrix of Fourier magnitude. (c): Similarity matrix of phase-weighted Fourier magnitude metric.

with the structural parameters could provide a better solution. Cunado *et al.* [103] proposed an early motion-based model to analyse the angular motion of the hip and thigh by means of a Fourier series. Fourier components of the motion of the upper leg were classified using k-nearest neighbourhood. They have shown that phase-weighted Fourier magnitude information gave better results rather than using only magnitude information.

Figure 2.10(a) shows the extracted hip rotation pattern for the sequences in a gait period of the particular subject using the proposed thigh model. The feature similarity matrix for each subject was generated using the Fourier magnitude metric and the phase-weighted Fourier magnitude metric. It can be seen from the Figures 2.10(b) and 2.10(c) that the phase-weighted Fourier magnitude metric demonstrates a greater separation between the mean signature of each subject than the Fourier magnitude metric. The algorithm was tested using image sequences captured indoors in an illumination controlled environment. A CCR of 100% was achieved when using the phase-weighted magnitude information and CCR of 80% when they used just the magnitude information.

Lee *et al.* [104] used ellipse fitting to the boundary pixels and extracted moment-based region features. Rather than taking the entire silhouettes, they divided the silhouette into regions and fitted ellipses based on the statistics on the region. Major and minor axes of the ellipse were defined using eigenvectors and eigenvalues based on the covariance matrix of each region. Orientation of the major axis, ratio of major and minor axis length, and position of the centroid from each region were used to form the feature vectors. the algorithm was tested on the CMU MoBo database [31] and an average CCR of 94% was achieved, when they used only 41 features. This result reduced to 91% when they used all 57 features.

Nixon and Yam [105] proposed a model that focuses on running rather than walking. They used a phase-weighted Fourier description with two approaches, based on the theory of coupled oscillators and the bio-mechanics of human locomotion. Both approaches derived a phase-weighted Fourier description of the gait signature by automated non-invasive models. They extracted the phase and magnitude components of the Fourier description of the thigh and lower leg rotation, measured within a gait cycle and used this as the gait signature. Though both models performed well, the analytical approach with a forced coupled oscillator motion model gave the better CCR of 85% for walking and 95% for a running person at rank 1.

Wang *et al.* [106] used static and dynamic fusion of vectors as a gait feature. They represented the pose changes of the segmented moving silhouettes as an associated sequence of complex vector configurations. Dynamic information such as joint angle trajectories of lower limbs coupled with static parameters of the model used with different combinations of rules improved the performance. Evaluation was performed using in-house data, consisting of 80 image sequences from 20 subjects. They used different kinds of rules, such as minimum, maximum and summation to model the gait features. Summation of static features and dynamic features gave the best CCR of 95% at rank 1.

Wagg *et al.* [107] proposed bulk motion and shape estimation guided by biomechanical analysis. They used anatomical data to generate shape models consistent with normal human body proportions and used mean gait data to create prototype gait motion models, which were adapted to fit individual subjects. Initially, bulk motion of the subject in the horizontal plane is estimated using a motion-compensated temporal accumulation algorithm [108]. Processed and filtered accumulations were evaluated using region-based and boundary-based template matching and the best matching person-shaped template was found, as illustrated in Figures 2.11(a) and 2.11(b).

The period of the gait is computed by measuring the sum edge strength within the outer region of the subjects' legs over time. The initial estimation of the shape is improved by computing a line Hough transform for each frame specific region. Following this, an improved estimated model is matched to image data as shown in Figure 2.11(c) to extract the joint parameters. This algorithm was tested using the SOTON database [66]. Performance of the algorithm was evaluated using CMS at rank 1 and achieved a CCR of 84% for indoor data and 64% for outdoor data.

Ziheng *et al.* [109] proposed a consistent Bayesian framework for introducing

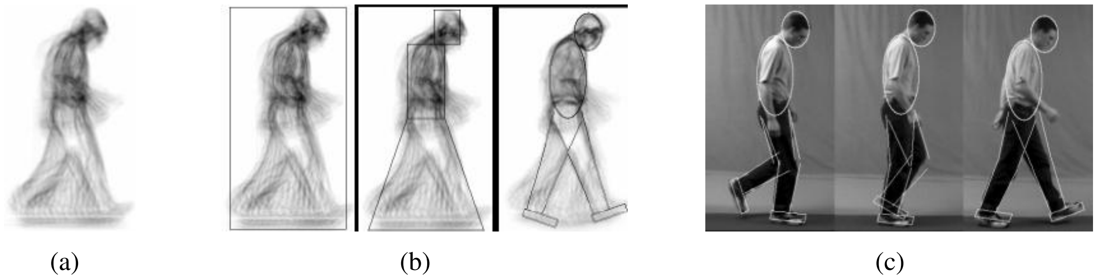


Figure 2.11: Algorithm Proposed by Wagg *et al.* [107] (a): Bulk motion estimation by temporal accumulation. (b): Shape estimation. (c): Model extraction process.

strong prior knowledge of human walking, and generated a simple human model for extracting gait. The strong prior knowledge is built from a simple articulated model having both time-invariant (static) and time-variant (dynamic) parameters. An HMM was used to fit the model to the extracted silhouettes. From the fitted model, the static parameters of the model were used directly and the dynamic parameters were compactly described by the coefficients of a Fourier series. They have tested the algorithm using the SOTON database [66] and achieved CMS of 61% in indoor data and 55% in outdoor data.

Model fitting in 3D space

Urtasun *et al.* [110] initially proposed a 3D tracking method that fitted 3D temporal motion models to the extracted silhouettes, which can help to recover motion parameters. An optical motion capture system is used to capture the stereo data for four subjects walking at different speeds. Motion tracking of these subjects was done using a PCA based tracker trained on low-resolution stereo data. The eigenvalues (weights) computed from the PCA tracker are used to recognize the people in the database and to characterize the motion of those who are not. The algorithm was evaluated over limited changes in view, changes in clothes and occlusion. It was shown that this method is robust to these variations.

Guoying *et al.* [111] proposed a complete 3D model with 10 joints (shoulder, elbow, hand, neck, head, knee, foot and hip) with 24 degrees of freedom. They used the video sequences from multiple cameras as input to set up the human model and motion was tracked by applying a local optimizing algorithm. Lengths of the key segments and motion trajectories of lower limbs were chosen as features for matching and recognition. Linear Time Normalization (LTN) and Dynamic Time Wrapping (DTW) were exploited to normalise the gait features. They tested the algorithm in the CMU

MoBo database [31] by selecting the slow walking dataset as the training images and the inclined walking dataset as the probe sequences, and achieved a CCR of 70%.

Tong *et al.* [112] proposed a 3D model represented by a convolution surface attached to articulated skeletons. The square of the Cauchy function [113] was used as the convolution kernel function to model the surface. Any arbitrary point on a surface was regarded as a line integral with unit line density. Then, this line was represented using a polynomial distribution function. This helps to counter the changes in polynomial parameters by using deformable shape. This 3D convolution surface is bridged with a 2D convolution curve using the curve correspondence theorem. New constraints defined for joint and skeletons improve their performance and enable them to work efficiently without self-occlusion. They tested their algorithm to recover walking human pose with 3D models with self-occlusion present in the video. They did not extend the model for gait feature extraction and recognition, however, the model would be suitable for this.

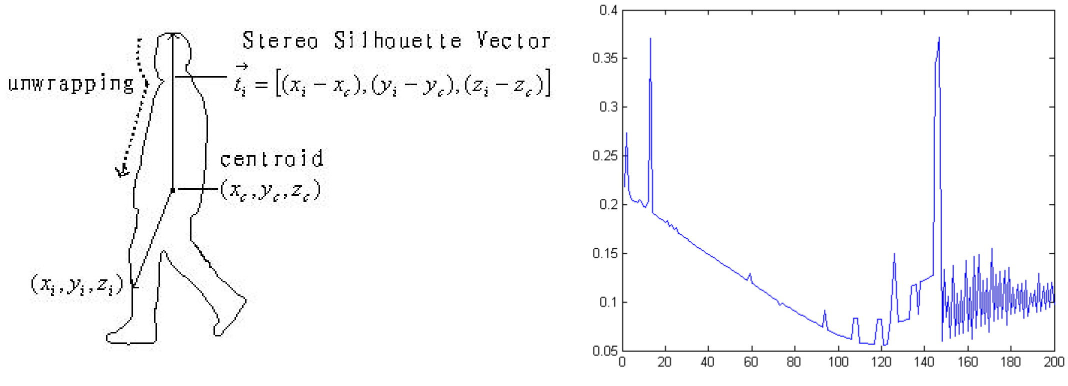
Haitao *et al.* [114] proposed a 3D method using Stereo Gait Features (SGF). They represented the moving 3D contour with a 1D-silhouette signal named the stereo silhouette vector (SSV), as represented in Figure 2.12. To compute the SSV, initially the centroid (x_c,y_c,z_c) of the silhouette was determined. Then the SSV (t_i) is defined as,

$$\vec{t_i} = \left[(x_i - x_c), (z_i - z_c), (z_i - z_c)\right], \tag{2.20}$$

where (x_i,y_i,z_i) is the pixel on the 3D silhouette, $i = 1, 2, ..., N$and N is total number of pixels. The 3D silhouette was unwrapped from top to bottom and all SSVs were computed to represent the 3D silhouettes as a 1D vector (F),

$$F = \left\{\vec{t_i}\right\}, i = 1, 2, ..., N. \tag{2.21}$$

The dimensionality of the data is reduced through PCA. They developed their own stereo database called PRLAB, which contains gait sequences, captured by a calibrated stereo camera pair. This database includes 14 different subjects with five sequences per subject. The algorithm tested in this database achieved a CCR of 92% with kNN and CCR of 70% with NN classifiers.



(a)

(b)

Figure 2.12: Example of stereo silhouette vector representing the moving 3D contour with a 1D silhouette signal [114]. (a): Unwrapping of Stereo Silhouette Vector (SSV) and (b): normalised L2-norm of 1D silhouette signal.

2.8 Post-processing of Extracted Features

Extracted features based on the methods described in the Section 2.7 require post processing tasks for better classification. Post processing of features includes reducing the feature dimensionality, identification of the most discriminative features and fusion of the selected features over time. The techniques used in existing research are outlined in this section.

2.8.1 Dimensionally Reduction - PCA and MDA

Principal component analysis (PCA) [115] and Multiple discriminant analysis (MDA)-[116] work in the least square sense. PCA represents the data in low dimensional space while maintaining the global Euclidean structure of the data in the high-dimensional space. However, MDA preserves discriminative information between data of different classes and tries to separate the data between two different classes by seeking better projection. Better data representation can be achieved by using both, as well as better class separability of data, while reducing the dimensionality. Most of the popular gait recognition techniques use both and have shown better recognition rate [117].

2.8.2 Compressive Sensing (CS)

Compressive sensing (CS) is a popular signal processing technique where a sparse signal is reconstructed from a small set of random projections [54]. In the recent literature, CS techniques have demonstrated promising results for signal compression and reconstruction. At the same time, the novel field of Compressive Sensing techniques has provided a new approach to the compression and reconstruction of signals at a rate significantly below that of Nyquist sampling, and has hence attracted much attention from signal processing researchers. The compactness of the CS representation makes it a very appealing technique also for the compression of time series in distributed pattern recognition applications. Apart from offering good compression capability, the relevance of CS to pattern recognition lies in its potential as a dimensionality reduction technique for series of sampled signals. Differently from techniques such as Principal Component Analysis, Linear Discriminant Analysis and many others, compressive sensing is not learned from a training set and therefore does not suffer from limited generalisation.

Currently there are a number of applications and researchers focusing on Compressive sensing for the following purposes in image processing areas.

1. As a dimensionality reduction technique.
2. Background subtraction.
3. Gradient domain processing.
4. Classification and detection.

In this thesis, compressive sensing is used to perform sparse representation-based classification that has been used as an effective classification method for face [52], action recognition [54] and facial expression recognition [118]. Comprehensive details of this classification method and the improvements that can be made for the optimised performance in gait context will be further illustrated in Chapter 6, as this is not explored in the gait recognition context in past literature.

2.8.3 Feature Selection Based on Analysis of Variance (ANOVA)

ANOVA is a standard technique for measuring the statistical significance of a set of independent variables in predicting a dependent variable. There are three types of ANOVA models:

1. Fixed-effect models assume that the data comes from normal populations that may differ only in their means.

2. Random-effect models assume that the data describes a hierarchy of different populations whose differences are constrained by the hierarchy.

3. Mixed-effect models describe the situations where both fixed and random effects are present.

A detailed description of ANOVA can be found in the book by Casella and Berger [119]. ANOVA has been used in gait recognition to explore image information important in silhouette-based gait recognition in [120].

2.8.4 Time Aggregation of Gait Features

There are a number of methods used to aggregate the temporal information of the gait features to compute the distance between the *probe* and the *gallery*. Popular methods used in the literature are summarised below.

Averaging

Averaging is the simplest and most compact way of summarising the gait feature vectors over time, based on the general assumption that gait features are normally distributed and hence can be represented by means and standard deviations. The average gait feature vector of a subject (for the total temporal frames or frames within the gait period) F is,

$$F = (mean_j(F_i), std_j(F_i)), \tag{2.22}$$

where j is the frame index, $i = 1, 2, ..., n$, and F_i is the i^{th} feature vector. Let $i = 1, 2, ..N$ and N is the number of extracted feature vectors, then the average representation of the particular subject has a $2 \times N$ dimensional feature vector.

The averaged feature set is very simple to compute and robust to noisy foreground silhouettes. The mean features describe the static parameters of the feature vector, while the standard deviation features roughly describe the dynamic changes of the feature vectors. Hence, averaging fuses the dynamic and static characteristics of the feature vectors. GEI [98] is the example feature vector that uses the averaging as a successful tool to accumulate dynamic information in a single row vector.

Time accumulated histogram

The accumulated histogram is a simplified and non-parametric feature modelling technique used in gait feature modelling. The only parameters that need to be globally assigned for the histogram of each feature are the number of bins and size of each bin. The similarity between two gait sequences is measured by comparing their complete sets of histograms. Boundaries of the histogram for each extracted feature (let's say N number of features) are computed as below,

$$\begin{aligned} \text{left edge}(f_i) &= \min_s(mean_t(f_i(s,t))) - \max_s(std_t(f_i(s,t))), \\ \text{right edge}(f_i) &= \max_s(mean_t(f_i(s,t))) + \max_s(std_t(f_i(s,t))). \end{aligned} \tag{2.23}$$

where s is the subject in the database and t is the frame index for each subject. The number of bins chosen is a trade-off between having good resolution in the histogram bins and maintaining a good estimation of the distribution, given the number of samples per sequence. In [121], 2D silhouettes are wrapped into an associated sequence of 1D signals to approximate temporal pattern of gait, and time accumulated histograms are used to represent the gait dynamics.

Decomposition of fourier harmonics

Since gait is analysed as periodic activity, harmonic components can be used to represent the temporal features [104,122]. Fourier decomposition to extract the fundamental and higher order harmonics of a given gait feature x with N number of temporal frames is achieved by,

$$X(k) = \sum_{j=1}^{N} x(j)e^{(-2\pi i)/N}(j-1)(k-1), \tag{2.24}$$

where $k = 1, 2, ..., N$ and $X(K)$ is the k^{th} complex Fourier form, representing the magnitude and phase. Intuitively, the magnitude measured at the fundamental frequency is a measure of the overall change undergone by the corresponding feature, and the relative phase between different time series is an indication of the time delay between the different features. Several gait recognition techniques particularly in model-based, multiple components of Fourier harmonics with amplitude and phase are used to represent the time series variation of gait features [104–106, 123].

Fundamental spectral decomposition

If the feature vectors of the particular gait cycle are used for Fourier decomposition, the gait fundamental spectral decomposition (F) for a particular subject for a particular gait period is,

$$FFT_i^1 = (|X_i(1)|, \text{phase}(X_i(1))), \tag{2.25}$$

where $i = 1, ..., N$ and N is the number of extracted feature vectors.

If the whole temporal frame of a particular subject is used for decomposition, then the fundamental frequency, Ω, is computed based on the highest peak in the power spectrum. Then, the temporal feature representation is defined as follows,

$$FFT_i^1 = (|X_i(\Omega)|, \text{phase}(X_i(\Omega))). \tag{2.26}$$

Similarly, the second and third (and so on) harmonics can also be computed. The optimal number of harmonics is dependent on the dynamics of the features. But, according to the review of the several approaches in the gait recognition field, fusion of the first three harmonics is shown as sufficient for the classification [123].

Direct sequence comparison using time normalisation techniques

Temporal gait features of a particular subject are better represented by the features of all the frames within a complete gait cycle, regardless of the number of frames required to represent the gait cycle. However, the length of the gait cycle differs for each subject, different walking conditions (i.e. surface, incline) and even within the same subject. This session-to-session and person-to-person variation causes the number of frames within the gait cycle to fluctuate. Therefore, before the comparison of gallery and probe temporal features, time normalisation techniques need to be applied to ensure that the recognition results are not biased by walking speed. Different techniques used in the literature for this purpose are outlined in the remainder of this section.

1. **Linear time normalisation (LTN)**
   Linear time normalisation computes the correspondence between frames in cycles of different length using a linear rule. In LTN, indexes of matching frames within gait cycles are computed as described in Figure 2.13, where the relationship between indices, x and y is $y = x * \frac{J-1}{I-1}$. Matching frame index (MFI) matrices for each r^{th} gallery cycle to t^{th} probe cycle are formed as below, Gallery to Probe indexes,

   $$(G_p^r) = \left\{G_p^r(1), G_p^r(2), G_p^t(i), ..., G_p^r(I)\right\}, \tag{2.27}$$

   probe to Gallery indexes,

   LTN is used in gait context [124] to match the different size of gait cycles to compute the same dimensional gait features for each individual.

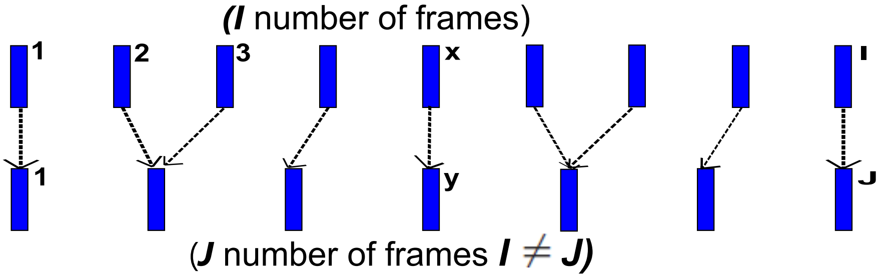


Figure 2.13: Illustration of computing LTN indices.

2. **Dynamic time warping (DTW)**

DTW is a dynamic programming tool used in speech recognition techniques to match the different sizes of speech signals of the same word. DTW also gave promising improvements to the performance for gait recognition [104]. The basic principle behind DTW is to allow a range of steps in time between samples (time frames of probe sequence, time frames in gallery sequence) and to find the path through that space that maximises the local match between the aligned time frames, subject to the constraints implicit in the allowable steps. For a given feature type where the probe and matching gallery are $G = g_1, g_2, ...g_m$ and $P = p_1, p_2, ...p_n$ respectively, where m is the number of frames in the gallery gait cycle and n is the number of frames in the probe gait cycle, DTW constructs an $m \times n$ matrix that contains the distances between samples g_i and p_i. Generally, the Euclidean distance is used for distance computation. The warping function, W, that maps one sequence to another is defined as,

$$W = \{w_1, w_2, ..., w_k, ..., w_K\}; \text{ where} \max(m, n) \leq K < (m + n - 1). \quad (2.28)$$

There are many ways to define the warping function, however, it must satisfy the following properties:

1. $w1 = (1, 1)$ and $w_k = (m, n)$: Ensures that the beginning and ending of the matching sequences are aligned.

2. Let the index of $w_k = (a, b)$, then $w_{k+1} = (a', b')$, where $0 \leq a' - a \leq 1$ and $0 \leq b' - b \leq 1$ to ensure the mapping of each warping step to the adjacent cells and the increasing warping with time.

3. The cost of the warping sequence G and P should be minimised.

The minimised cost between warping sequences as shown in Equation 2.29 can be directly used as a similarity measure for classification purposes.

$$DTW(G, P) = \min \left(\frac{\sqrt{\sum_{k=1}^{K} w_k}}{K} \right). \tag{2.29}$$

Compared to the direct sequential matching using a moving window [67], LTN and DTW, both improve the performance and computational expense. However, Boulgouris *et al.* [124] mentioned that they have found that LTN performs better for gait recognition approaches in contrast to DTW, which is recognised as the better method in speech recognition.

2.9 Classification Techniques

Classification techniques are proposed to recognise test objects by matching the extracted discriminating features on the already known objects. In this section, popular classification techniques used in gait recognition context are illustrated.

2.9.1 k Nearest Neighbour (kNN)

kNN is a simple, instance-based but powerful classification method commonly used in gait recognition [125]. The training phase for kNN consists of simply storing all known subjects and their subject labels. In the testing phase, initially, distances for a particular probe to the known subjects in the gallery set are computed using one of the well-known distance measures shown below, Euclidean distance is commonly used in gait recognition and it's distance can either be computed using the normalised features, or weighted by the significance of each dimension. Normalisation of each dimension in the feature vector is done by subtracting out the mean of that dimension and then dividing by the corresponding standard deviation. Features also can be weighted based on their significance towards distinguishing individuals. Weights of the particular dimension can be computed by feature selection techniques such as ANOVA (Section 2.8.3). If the features are independent and the influence towards the classification is not sure, then each dimension is weighted equally based on the assumption that all dimensions are equally significant. Euclidean distance, d between two gait features g and p with

N dimensions is,

$$d = \sqrt{(g-p)WC(g-p)^T}, \tag{2.30}$$

where C is the covariance matrix of the gait features and W is a diagonal matrix containing weights for each dimension.

Computed distances for each probe and gallery are ordered from lower value to higher and the first k subjects in the gallery will be selected. The most frequent subject in the k nearest neighbour is returned as the matching subject. kNN is simple to implement, robust with regard to the search space and the classifier can be updated in real time at very little cost as new instances with known classes are presented. There are only two parameters needing to be tuned for this algorithm: distance metric and k. kNN is sensitive to noisy or irrelevant attributes, which can result in less meaningful distance measures. Scaling and/or feature selections are typically used in combination with kNN to mitigate this issue. Most of the gait recognition approaches [125–127] in literature used kNN for their classification.

2.9.2 Support Vector Machines (SVM)

Support Vector Machines are based on the concept of decision planes that define decision boundaries [128]. A decision plane is one that separates between a set of objects having different class memberships. The idea behind SVMs is to make use of a mapping function ϕ that transforms data in input space to data in feature space in such a way as to render a problem linearly separable. The SVM then automatically discovers the optimal separating hyper plane (which, when mapped back into input space via ϕ, can be a complex decision surface). Given a training set of instance-label pairs (x_i, y_i), $i = 1, ..., T$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$, different types of kernels as listed below can be used to map the data for SVM classification.

- Linear: $K(x_i, x_j) = x_i^T x_j$.
- Polynomial: $K(x_i, x_j) = (\gamma x_i^T + r)^d, \gamma > 0$.
- Radial Basis Function (RBF) : $K(x_i, x_j) = \exp(-\gamma \left\|x_i - x_j\right\|^2), \gamma > 0$.
- Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$.

where γ, r, and d, are kernel parameters. The RBF is by far the most popular choice of kernel type used in support vector machines. This is mainly because of its localised and finite responses across the entire range of the real x-axis. Only a few types of gait

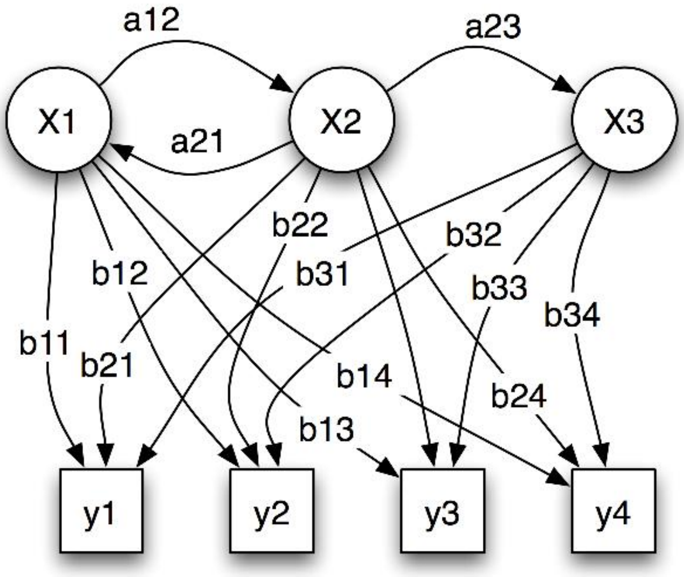


Figure 2.14: Probabilistic parameters of a Hidden Markov Model [131]. $x1, x2, x3$ are the hidden stages and $y1, y2, y3, y4$ are the observation. HMM is defined by (a) the transition probabilities and (b) the observation probabilities.

recognition research [129, 130] use support vector machines as a classifier, since they need model training to perform the classifier.

2.9.3 Hidden Markov Model (HMM)

A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states [131–133].

Unlike a general Markov model, in a hidden Markov model, the state is not directly known, but output, dependent on the state, is known. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states. Since the gait features are based on a temporal pattern, HMMs can be used as it has shown strong results in the similar temporal processes such as speech, handwriting, gesture recognition etc. Figure 2.14 shows an example HMM model with its probabilistic parameters. As shown in the Figure 2.14, the hidden states are $x1, x2, x3$ and the observations are $y1, y2, y3, y4$.

Then, there are three probabilistic parameters defined to describe the HMM:

1. Transition probabilities: $a_{ij} = P(x_i|x_j)$, i.e, probability for the next state to be x_i, if the given state is x_j.

2. Observation probability: $b_i(y_j) = P(y_j|x_i)$, i.e, probability to get the observation y_j if the current state is x_i.

3. Initial state probability: $\Pi_i = P(x_i)$, i.e, probability to the initial state to be x_i.

Typically an HMM can be defined as $M = (A, B, \Pi)$, where A is the matrix of transition probabilities, B is the matrix of observation probabilities and Π is the matrix of Initial state probabilities. Then, based on these probabilistic matrices, there are three problems that need to be solved; learning, decoding and evaluation.

The learning problem finds the most likely set of state transition and output probabilities for a given output sequence. A solution to solve this problem is given by the Baum-Welch algorithm [134]. In the decoding problem, the most likely sequence of (hidden) states which could have generated a given output sequence are found for the given model parameters. This is solved by the Viterbi algorithm [135]. Finally, the evaluation problem computes the probability of a particular output sequence for a given model. Forward and backward algorithms explained in [136] are used to solve this problem. Kale *et al.* [137] showed improved recognition performances using HMM-based classifier in gait context using frame exemplar distance gait features.

2.9.4 Summary

HMMs gather the temporal information and can give better performance for classification tasks. However, it requires plenty of training data to train the HMM model and it is computationally expensive. Similarly, SVMs are also complex, and need a lot of data to train the model and need a suitable kernel function for optimum results. But, for different features and conditions, a single general kernel is not appropriate. kNN is simple and performs reasonably well with less computational cost and can be achieved in near-to-real time, but it is sensitive to noisy or irrelevant attributes. Therefore, choice of classification method depends on the availability of data and the type of application.

2.10 Summary

Based on the extensive study on gait recognition literature, the following conclusions are made.

1. Higher performance in the range of 95-100% recognition rate in similar laboratory conditions is achieved in recent gait recognition research [44]. This is an evidence to use gait for distinguishing individuals. In particularly, it shows the potentials of gait to be used with the other biometric for robust human recognition. Research on gait features for human recognition is more focused on using them at unique application scenarios where recognition needs to be performed without alerting or interacting with the subject.

2. Most of the existing methods are performed on side-view as it better represents walking variations/dynamics [138]. However, the recognition performances of these approaches heavily degrade when an individual changes his walking direction or the camera view angle changes. Since these methods are based on 2D images, extracting the walking dynamics on front-to-parallel view of the camera is the main limitation of these methods. Particularly poor performances are achieved in frontal view due to this.

3. Appearance-Based methods achieved higher recognition rate in similar conditions compared to model-based approaches as they represent more richer gait features with structural and dynamic gait information. Model-based approaches struggle to perform better as they need higher resolution image data for better robust model-fitting, which cannot be obtained when subjects are walking at a distance. Model-based approaches also need higher computational resources for complex matching and searching. However, clean-data model-based approaches have the potential to perform consistently, regardless of the other appearance changes such as view, clothing or carrying things.

4. Most of the recent gait recognition approaches are based on appearance-based methods because of their higher recognition rate and simplicity. Particularly, they focus on providing solutions to apply to real world conditions, where appearance changes need to be addressed. Within the several gait challenging conditions, most common appearance changes that could occur commonly in real-world applications are clothing, carrying things and view.

5. Little research has been performed on 3D data, though most of it focuses on model-based techniques and is evaluated on the in-house developed (not publicly available), multi-view databases with a few number of subjects.

6. The current trend of biometrics-based application is to use them on biometric smart gates that can be implemented to provide office security up to border security in the airports [139]. Face and iris are already used in these applications, however, they still need human interaction. Gait features enable this to be performed in an non-intrusive way by fusing the frontal gait with these near-field biometrics. However, sparse research only is conducted on frontal gait, as gait dynamics cannot be accurately captured in frontal 2D view.

7. Forensic applications require gait recognition, which needs to be performed in arbitrary view angle only using the image data from surveillance cameras. Although, there are many techniques with the view transformational model are proposed, poor recognition rate in the range of 20-30% is achieved when the view angle or walking directions changes with more than 30° [140].

The research gabs discovered and current motivations provide the pathways and directions of the research conducted in this thesis. The remaining chapters of this thesis illustrate the contributions and outcome of the research that addresses the essential parts of these research gaps.

Chapter 3

Gait Recognition Framework

3.1 Introduction

The following chapter describes the gait recognition framework associated with the baseline implementation and initial steps taken to enhance the baseline for robust recognition performance. It also outlines the databases gathered and testing protocols that were used to evaluate the proposed algorithms.

Based on the literature review (Chapter 2), it has been noted that appearance-based gait recognition methods show significantly better performance compared to model-based techniques, particularly in a controlled environment, when there are no major changes in subject's appearance. Therefore, most of our research works focus on appearance-based techniques especially, to maintain or improve this robust performance towards real world challenges.

To understand the challenges in appearance-based gait recognition and to evaluate our proposals throughout the research, the GEI-based feature extraction technique proposed by Han and Bhanu [98] is chosen as the baseline, because of its simplicity and higher recognition performance. Since this method works in 2D, it shows its best recognition performance in side-view, due to the better appearance representation of gait dynamics. Therefore, recognition rate of GEI implementation in side-view is taken as a promising benchmark for comparing the outcomes of this thesis.

The first part of this chapter outlines the databases and testing protocols used in our evaluations. Following that, each element in the gait recognition framework and its implementation in the baseline context is explained. A generalised form of gait recognition framework and functionality of its elements is outlined in Section 2.5 in Chapter 2. Figure 3.1 narrows that down and shows the basics steps in GEI-based gait

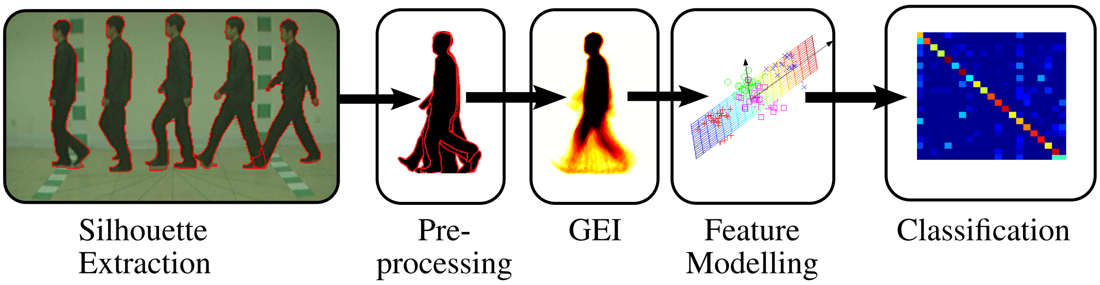


Figure 3.1: GEI based gait recognition framework. Silhouettes are extracted, pre-processed and GEI-based gait features are extracted. The extracted features are modelled and classification is done on these better discriminated features.

recognition system. As a preliminary step, silhouettes are extracted, normalised and registered in-order to compute the temporal image template, the GEI. The GEI features are then transformed to discriminative domain for better and computationally efficient classification.

Throughout the detailed implementation of our baseline, each element of the framework is explored towards better gait recognition performance, and particularly, extensive analysis is carried out in pre-processing tasks to make clear their influence towards recognition performance. Conclusions resulting from the analysis on segmentation and pre-processing tasks are followed in the analysis on the rest of the elements.

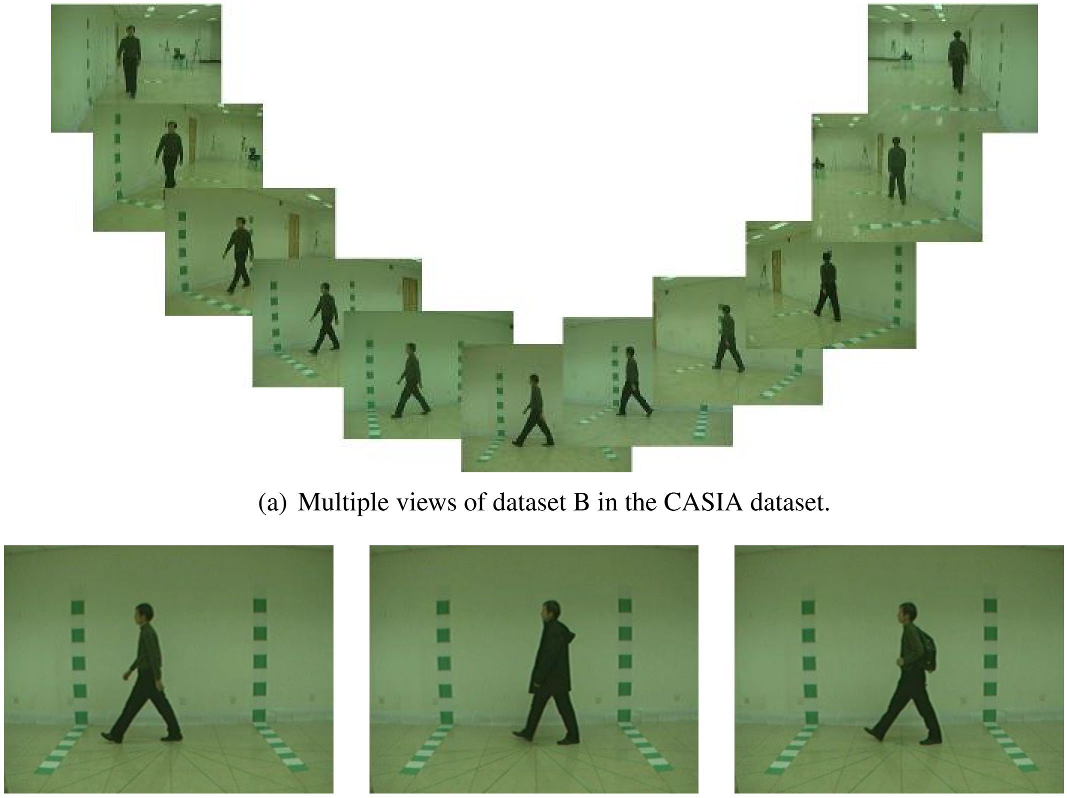
The remainder of this chapter is organised as follows. Section 3.2 explains the obtained gait databases and preparations of testing protocols to evaluate the gait recognition algorithm in them. Section 3.4.1 discusses the improved silhouette segmentation and image registration. Section 3.3 outlines the baseline feature extraction process using GEI-based features and Section 3.4 explains enhanced analysis made to study the limitations in the baseline approaches and improve the recognition performance, while Section 3.5 concludes the chapter by summarising the outcomes.

3.2 Organisation of Test Dataset

To evaluate our algorithms following most popular gait datasets, CASIA dataset B [55], OULP [46] and CMU MoBo [31] are obtained and organised according to the common testing protocols distributed with them. The remaining part of this section describes each of these datasets in terms of number of subjects, available challenging conditions to explore and corresponding testing protocols.

3.2.1 CASIA Dataset

The CASIA dataset [55] is a large population database with multiple views and different challenging conditions. The dataset contains 124 subjects, with three different walking conditions: normal walk (*nw*), bag (*bg*) and clothing (*cl*). There are multiple numbers of recorded sequences provided for each subject and each condition. For *nw*, six sequences exist for each subject, while there are two for *bg* and *cl*. Figure 3.2 illustrates the available varieties in this dataset.



(a) Multiple views of dataset B in the CASIA dataset.

(b) normal walk (*nw*)

(c) with coat (*cl*)

(d) carrying bag (*bg*)

Figure 3.2: Example images from the CASIA dataset B. (a) Walking sequence from eleven camera views, (b) normal walking (*nw*), (c) walking with different clothing (*cl*) and (d) walking with a bag ((bg).

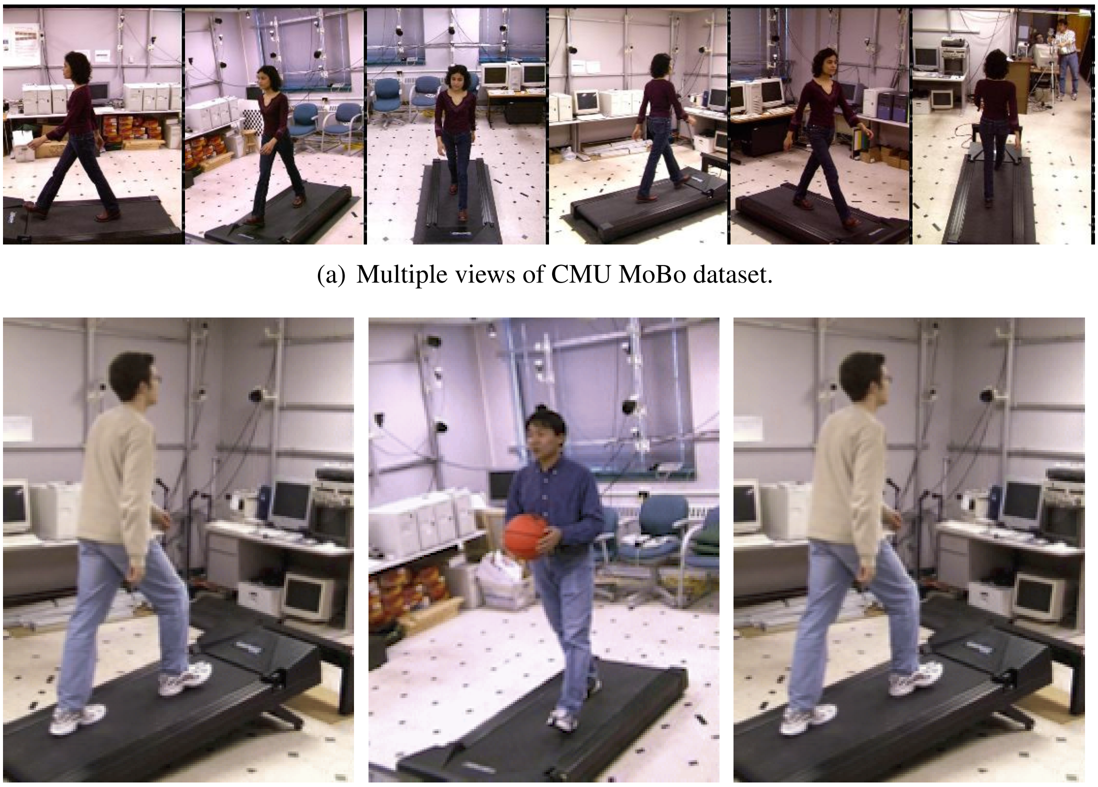
The experiments on this dataset follow the evaluation outlined in [141]. The intra-class experiment is performed on the *nw* sequences, with four allocated to the gallery and two to the probe. In inter-class tests, again four cycles from *nw* are used as a gallery, while the two sequences in each of the other classes make up the probe in their individual experiments. Table 3.1 summarises the experiments for the CASIA database.

Experiment	Gallery	Cycles	Probe	Cycles
nw	*nw*	124×4	*nw*	124×4
nw-bg	*nw*	124×4	*bg*	124×2
nw-cl	*nw*	124×4	*cl*	124×2

Table 3.1: Test cases on the CASIA database.

3.2.2 CMU MoBo Dataset

The CMU MoBo gait dataset [31] is the only available multi-view database, where subjects are captured from six cameras simultaneously, as they walk on an indoor treadmill. The database consists of 25 subjects in four different walking conditions: slow walk, fast walk, walking while carrying a ball and walking on an inclined surface.



(a) Multiple views of CMU MoBo dataset.

(b) Slow Walk (*sw*)

(c) Carrying a Ball *bl*

(d) Incline *sf*

Figure 3.3: Example images from the CMU MoBo dataset (Multi-view images with different conditions).

To enable cleaned segmentation, the backgrounds of each view for slow walk, fast walk and walking while carrying are distributed with the database, while not provided for inclined surface data. Due to the improper distribution of incline data, this is not included in the experiments carried out in this thesis. Each subject is recorded for 340 frames of duration in each view, for each walking condition, using 30 frames per second frame rate. Each gait condition in this database is recorded under the following criteria:

- **Slow Walk**
  The speed of the treadmill is maintained to make comfortable walking for each individual and average speed of $3.32\,\mathrm{km\,h^{-1}}$ is recorded.

- **Fast Walk**
  Here, each individual is directed to walk fast with comfort. Average walking speed of $4.54\,\mathrm{km\,h^{-1}}$ is recorded.

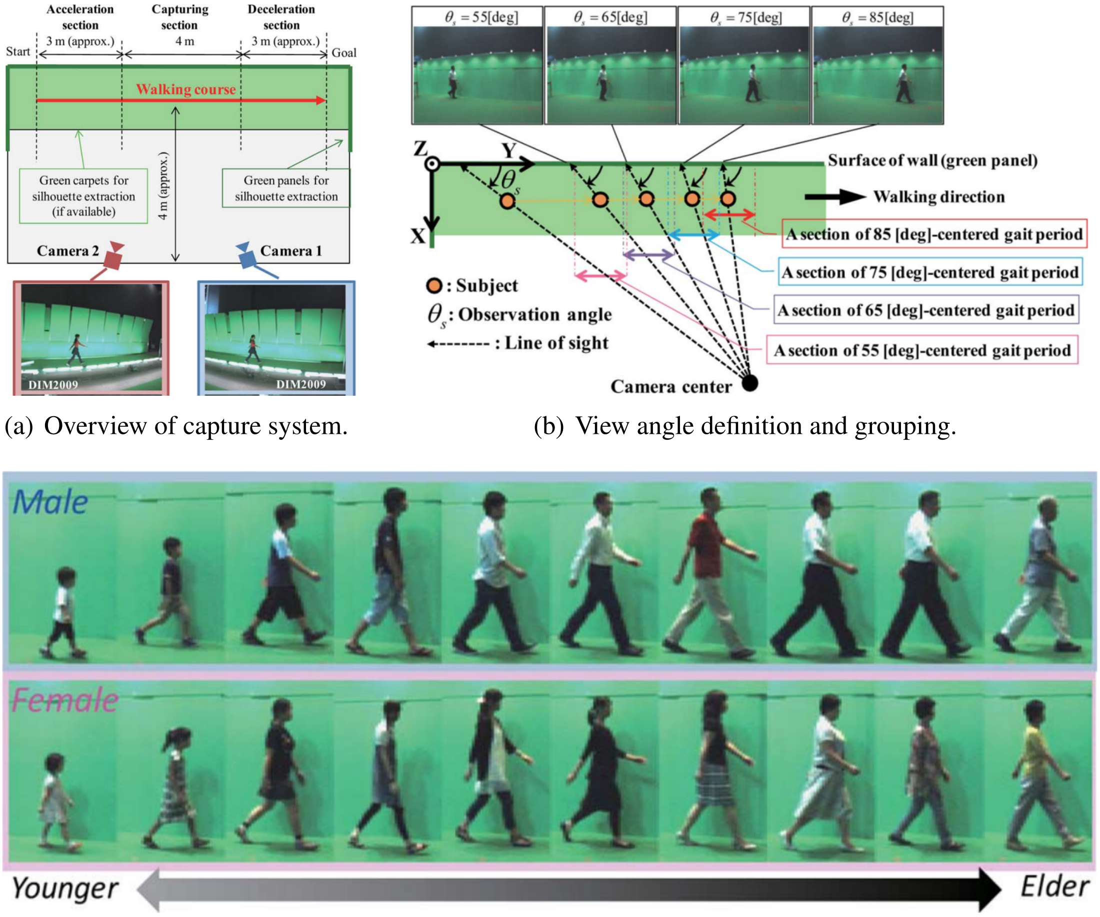
- **Incline walk**
  The treadmill was set to the maximum incline of $15°$. The speed was adjusted to $3.15\,\mathrm{km\,h^{-1}}$ to enable comfortable walking.

- **Walking with a ball**
  In this case, subjects are directed to take a comfortable walk, while holding a ball in front of their body. The objective behind this sequence is to immobilise the arms and analyse how this affects their gait pattern. Average walking speed of $3.28\,\mathrm{km\,h^{-1}}$ is recorded.

Experiment	Gallery	Cycles	Probe	Cycles
sw	Slow Walk	25×3	Slow Walk	$25 \times 3 \sim 4$
fw	Fast Walk	25×3	Fast Walk	$25 \times 3 \sim 4$
bl	Ball	25×3	Ball	$25 \times 3 \sim 4$
sw-fw	Slow Walk	$25 \times 6 \sim 7$	Fast Walk	$25 \times 6 \sim 7$
sw-bl	Slow Walk	$25 \times 6 \sim 7$	Ball	$25 \times 6 \sim 7$
bl-fw	Ball	$25 \times 6 \sim 7$	Fast Walk	$25 \times 6 \sim 7$

Table 3.2: Test cases on the CMU MoBo database.

The following commonly used testing protocol, similar to the one in [88], is used in evaluations of algorithms on this database. By exploring the stride length in side-view, around 6-7 gait cycles are segmented for each subject, in each condition. For intra-class test cases, the first half of these gait cycles are used for training and the remaining are used for testing. In the inter-class case, all the available gait cycles are used as gallery and probe. Table 3.2 shows the intra and inter-class test cases available for the CMU MoBo database.



(a) Overview of capture system.

(b) View angle definition and grouping.

(c) Examples from different age group and gender

Figure 3.4: Capturing system of the OULP database [46].

3.2.3 OULPC1V1-A dataset of the OU-ISIR Gait database

OULPC1V1-A dataset of the OU-ISIR Gait database (OULP) [46] consists of silhouettes from a maximum of 3961 subjects, which are grouped based on the captured view angle (55, 65, 75, 85, All) from two cameras, as shown in Figure 3.4.

The database consists of subjects of age range from 1 to 94 years and with equal ratio of female and male candidates, as shown in Figure 3.4(c). Each subject has two sequences, and from them, one centred gait cycle for the particular view is selected. Cleaned segmented silhouettes are distributed as a part of this database; the first sequence is used as gallery and the other one is used as probe, to form the intra-class test case as outlined in [46]. For the baseline evaluation, gait cycles from near-profile view (*A-85*) and gait cycles that cover all the angles (*A-All*) will be considered.

3.3 Baseline Implementation

In this section, the implementation of gait recognition framework is outlined using a simple and well-performing gait feature extraction method GEI [98]. GEI is chosen, due to the potential recognition performance with less computational complexity. GEI has strong opportunities for future enhancements, and most of the recent gait research is more focused on optimising these GEI features for different gait-based research directions. Silhouettes distributed with datasets are used in the baseline, to make sure there is a fair comparison with the methods used in literature. After silhouettes are extracted from video footage, GEI features are extracted. Conventional PCA and MDA are used to extract discriminating features from the GEI, and similarity distances are computed in the learned basis domain.

As a first step in the baseline implementation, silhouettes are centrally aligned, based on centre of mass. For this purpose, the region of interest (ROI) is extracted by defining the bounding box on the existing pixels on the binary image. Then, computed ROI is resized, using nearest-neighbour interpolation for the fixed height. This resized ROI is aligned to the horizontal centre of mass (mean value of the horizontal coordinate of existing pixels on ROI) and cropped or padded with zeros, to maintain the fixed width.

3.3.1 Gait Cycle Segmentation

Gait uses a combination of physical and behavioural characteristics of an individual as an identity. Hence, it is required to gather temporal patterns to represent the dynamics. However, the psychological and medical studies on gait, state that gait is periodic and therefore, it is repetitive after each gait period [48]. Therefore, gait feature of an individual can be represented by features extracted within a gait period.

Gait period can be estimated by analysing the periodic nature of parameters extracted from silhouettes. Stride length, pixel solidity ratio, normalised height before alignment and model parameters are mostly used in literature to segment the temporal frames into a gait period. However, stride-length is most commonly used in side-view, as it is the view better representing the periodicity that directly attributes to heel strikes.

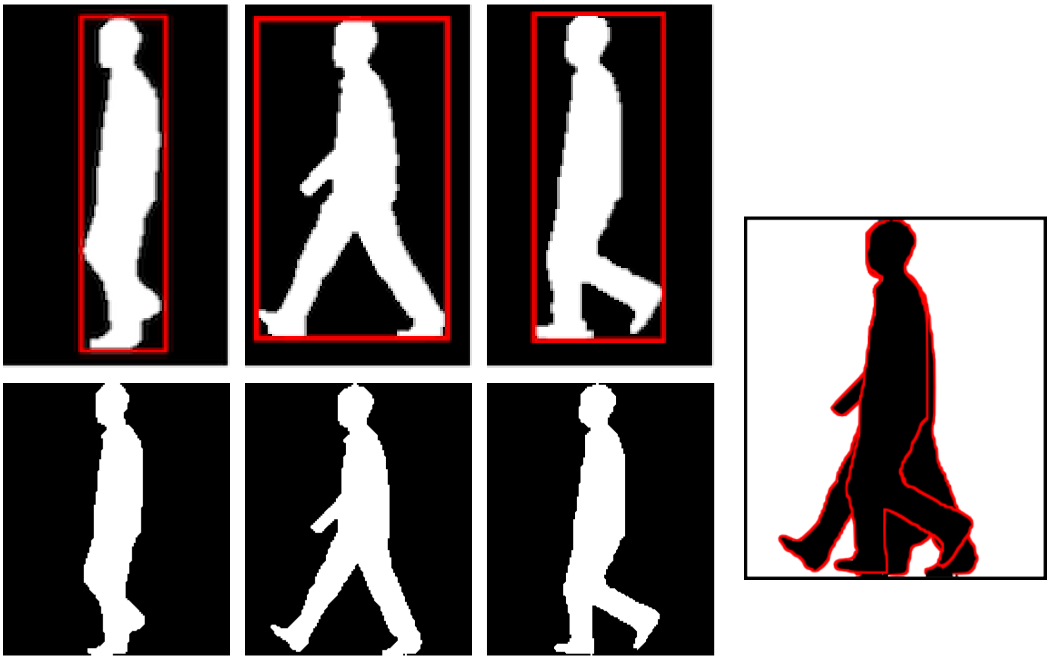


Figure 3.5: Pre-processing of extracted silhouettes.

Based on that, stride length is used to segment the gait cycles in side-view experiments in our research. From the aligned silhouettes, width of the region of interest area (where silhouette pixels exists) is computed and the signal is smoothed, using a median filter as shown in Figure 3.6. Peaks in the smoothed width profile are detected as they represent the extreme strike. The distance between every second peak is computed as gait cycles, as shown in Figure 3.7.

3.3.2 Feature Extraction

As explained in Section 3.1, gait features are extracted by computing the temporal image template, GEI, following the method proposed by Han and Bhanu [98].

A simple and well-performing gait feature representation, the GEI, has been chosen a feature extraction method in baseline implementation. Inspired by the motion history images (MHI) and the motion energy images (MEI) [142] proposed for action recognition, gait energy images [98] were developed for use in gait recognition.

GEI represents the gait features in multiple silhouettes of a person over a gait cycle in a single image frame. This forms a compact representation of a person's spatial occupancy over a gait cycle, encoding information about their gait characteristics, as well as appearance, allowing identification to be performed.

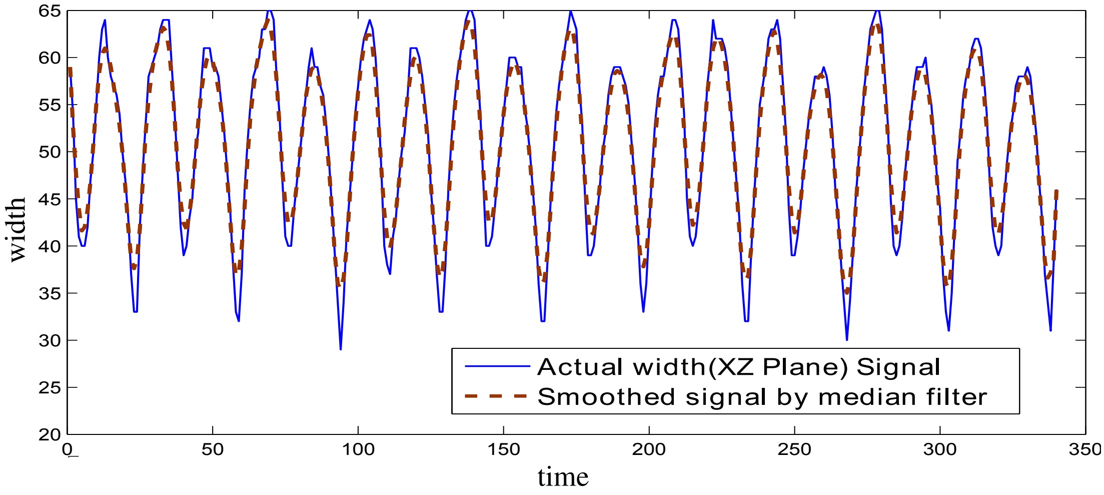


Figure 3.6: Temporal pattern of the width of the silhouettes and median filter smoothing (Example using 350 frames of an individual from the CMU MoBo database).

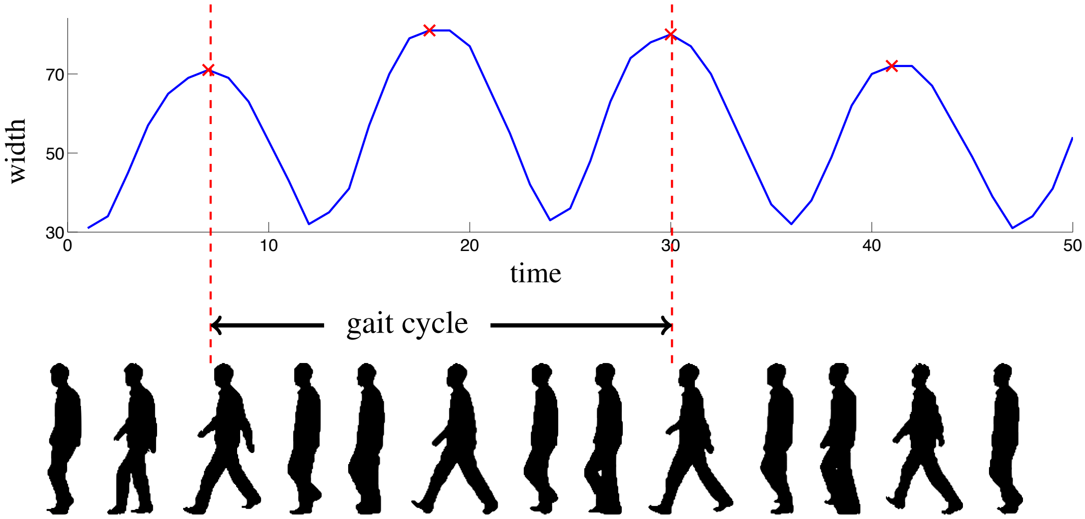


Figure 3.7: Width profile corresponding to strikes and gait cycle segmentation (Example using 50 frames of walking individual from CASIA database.)

To compute the GEI, pre-processed silhouettes are averaged over gait cycle as follows,

$$\mathrm{GEI}(k) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{I}(t), \tag{3.1}$$

where n is the number of frames in the k^{th} gait cycle and I is the normalised silhouettes. Example silhouettes and computed GEI from the CASIA dataset are shown in

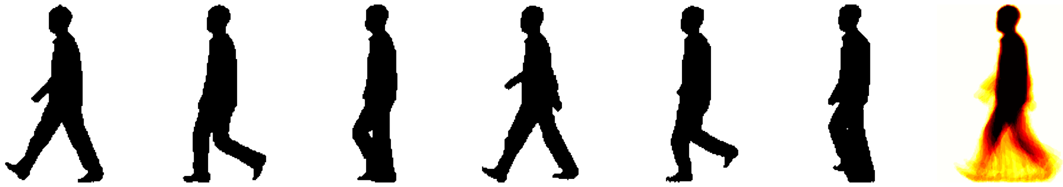


Figure 3.8: Gait energy image. Selected silhouettes from an example gait cycle along with its corresponding GEI.

Figure 3.8.

To perform post-processing tasks, input feature image, $GEI \in \mathbb{R}^{r \times c}$, is wrapped and concatenated in single column vector, $X \in \mathbb{R}^{d}$ where $d = m \times n$.

3.3.3 Feature Modelling

It is computationally expensive to use the higher dimensional input feature X directly, to perform the classification. It is also important to extract the most dominant feature components that contribute to discriminate the subjects, while discarding the unwanted components. This section outlines the implementation process of feature modelling techniques with basis learning, to transform the input feature space to a more discriminative feature domain.

Multiple discriminant analysis (MDA) [116] is an effective discriminative dimensionality reduction technique, commonly used in computer vision applications. MDA learns a basis on the training data to transfer the feature space into a most discriminative domain. However to make it robust, input feature space is required to be given, with the most energetic components. This can reduce the influence of non-energetic noise components in the learning process in MDA, at the same time as it inputs reduced dimensional space to MDA for efficient learning. For this purpose, principal component analysis (PCA) is performed on the input feature space to extract the most energetic components, regardless of individual subject labels.

Dimensionality reduction using PCA

To learn the PCA basis, dictionary, $D = X_1, X_2, ..., X_n \in \mathbb{R}^{d \times n}$, is formed from the training data, where n is the number of feature images on training data.

To learn the basis that can retain more energetic components on the feature space, PCA attempts to minimise the error function,

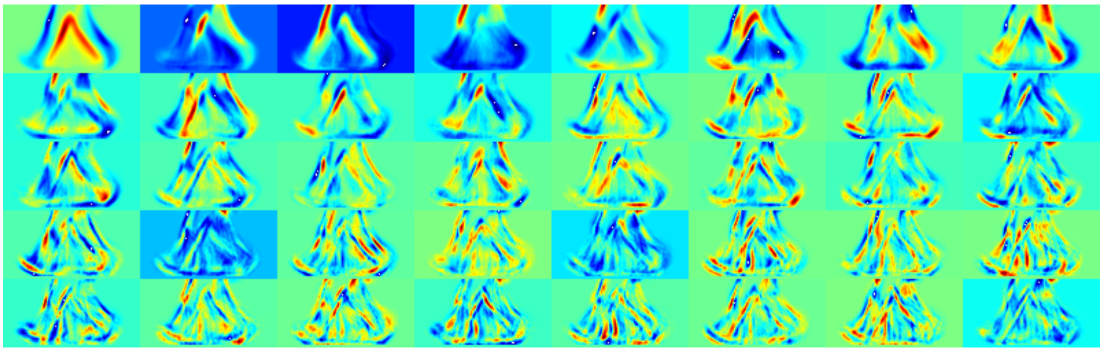
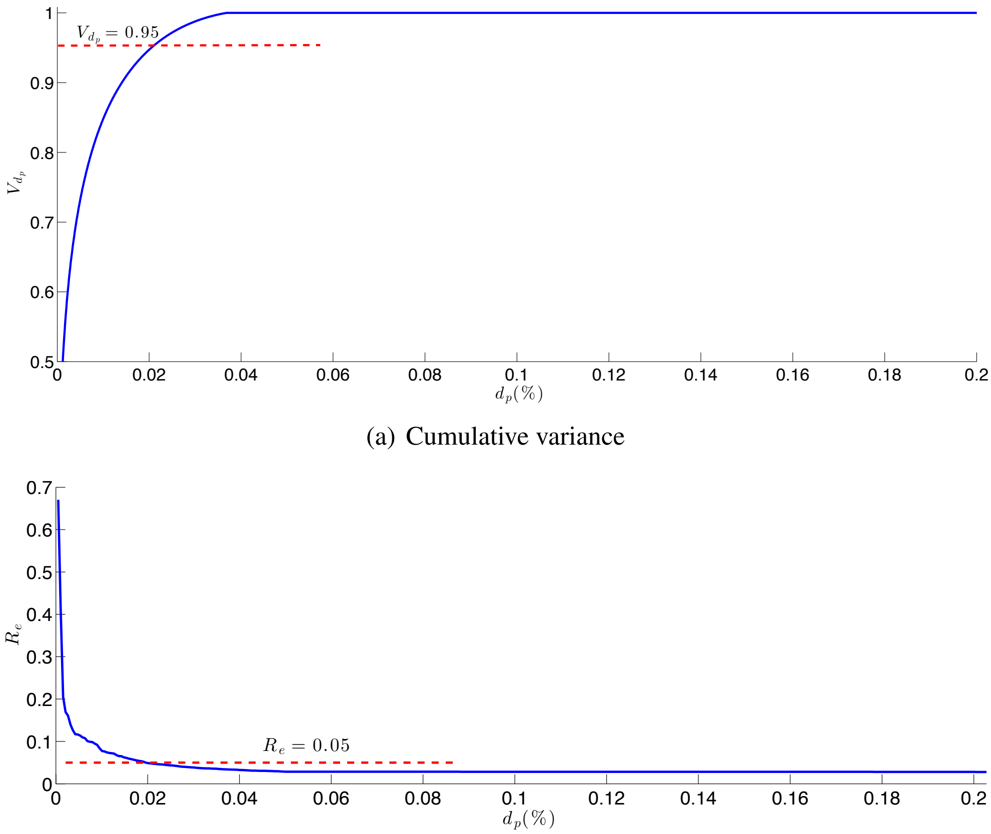


Figure 3.9: Example of PCA components from the learned basis. First 40 components with higher variance are shown.



(a) Cumulative variance

(b) Reconstruction error

Figure 3.10: Change of cumulative variance and reconstruction error with ratio of principal components. Using examples from GEI-based feature space.

$$J_{d_p} = \sum_{k=1}^{n} \left\| (m + \sum_{j=1}^{d_p} a_{kj} e_j) - X_k \right\|^2, \tag{3.2}$$

where $d_p < d$ is the number of principal components that can be chosen for better representation of feature space with least reconstruction error, m is the mean, defined as $m = \frac{1}{n} \sum_{k=1}^{n} X_k$. $e_1, e_2, ..., e_{d_p}$ are set of orthogonal unit vectors. X_k is projected into this space to form a. The error, J_{d_p}, is minimised when $e_1, e_2, ..., e_{d_p}$ are eigenvector with the largest eigenvalues. Figure 3.9 illustrates the first 40 principal components computed from a example GEI feature.The projected d_p dimensional feature vector after the PCA transformation is obtained from X_k as follows,

$$\begin{gathered} Y_k = T_{pca}^T \times X_k = [a_1, a_2, ..., a_{d_p}]^T = [e_1, e_2, ..., e_{d_p}]^T X_k \in \mathbb{R}\shortmid, \\ \text{where, } k = 1, 2, ..., n. \end{gathered} \tag{3.3}$$

Selection of principal components from PCA basis

Selecting the number of dimensions required for an effective PCA-based dimensionality reduction is crucial, as it needs to preserve the principal components corresponding to higher variance that can result in lower reconstruction error.

Variance of principal component can be computed, based on the corresponding eigenvalues. Cumulative variance level (V_{d_p}) for the first d_p number of principal component is calculated as below,

$$V_{d_p} = \frac{\sum^{d_p} l_i}{\sum^{d} l_i}, \tag{3.4}$$

where l_i is $i^t h$ variance from a latent matrix that is ordered from higher to lower. Figure 3.10(a) shows the latent variance, l, with increasing number of principal components. It can be noticed that more than 95% of cumulative variance is represented in less than 2% of principal components. To understand the actual impact of major principal components on GEI image, the GEI is transformed and reconstructed, using the computed basis with a different number of principal components as below,

$$\hat{X_\gamma} = T_{pca} Y_k + \mu(D) \in \mathbb{R}^{d_p}, \tag{3.5}$$

where $\mu(D)$ is mean of the input features in the training dictionary, D. Based on this, the averaged reconstruction error, $R_e(d_p)$, of the training features for the dimension of

dp, can be computed as below,

$$R_e(d_p) = \sum_{k=1}^{n} \frac{\left\| (m + \sum_{j=1}^{d_p} a_{kj} e_j) - X_k \right\|^2}{\|X_k\|^2}, \tag{3.6}$$

Figure 3.10(b) shows reconstruction error with the different percentage level of principal components. It has also been noticed that less than 5% error is occurred with only the use of 2% of the principal components. Based on the latent and reconstruction error analyses on the principal components, d_p is chosen, as that corresponds to less than 5% reconstruction error as the optimum for recognition performance and computational efficiency.

Transformation to discriminant domain using MDA

These new features better represent the original features in fewer dimensions with minimised distortion error. d_p is chosen to preserve as an optimum balance between reduction in dimensionality and reconstruction error.

However, all of these features might not be significant towards discriminating individuals. By learning an MDA basis on these input features, the input features can be transformed to the discriminated domain that represents the individual's uniqueness. Suppose n d^p dimensional feature vectors $Y_1, Y_2, ..., Y_n$ correspond to c classes. MDA finds the transformation matrix to maximise the ratio of between-subject (inter-subject) to within-subject (intra-subject) variance J_{d_m} as follows,

$$J_{d_m} = \frac{|\tilde{S_B}|}{|\tilde{S_W}|} = \frac{|W^T S_B W|}{|W^T S_W W|}. \tag{3.7}$$

S_W is the within-subject scatter matrix, defined as $S_W = \sum_{i=1}^{c} S_i$. S_i can be computed as $S_i = \sum y \in D_i (y - m_p)(y - m_p)^T$, where $m_i = \frac{1}{n_i} \sum y \in D_i y$. D_i is set of training templates that belong to the i^{th} subject. n_i is the number of templates in D_i. S_B is between-subject scatter matrix, computed as $S_B = \sum_{i=1}^{c} n_i (m_i - m_p)((m_i - m_p)^T$, where m_p is the mean of the PCA transformed feature vectors, defined as $m_p = \sum_{i=1}^{n} y_i$. J_{d^m} is maximised when the columns of W are the generalised eigenvectors of S_W and S_B that corresponds to the largest eigen values and satisfy equation below,

$$S_B W_i = \lambda S_W w_i. \tag{3.8}$$

The MDA transformation is applied to PCA projected vectors $(y_1, y2, ..., y_k)$ as follows,

$$Z_k = T_{mda}Y_k = [w_1, w_2, ..., w_{d_m}]^T, k = 1, 2, ..., n. \tag{3.9}$$

Selection of d_m depends on the number of subjects. It is obvious d_m cannot be greater than the number of classes. Therefore, d^m is defined as $d_m = c - 1$. The Computed PCA transformation (T_{pca}) and MDA transformation (T_{mda}) from the training sequences are used to produce the dimensionally reduced transformed vector (Z_k) of the testing feature (X_k) as follows,

$$[Z_k]_{d_m \times 1} = [T_{mda}]_{d_m, d_p} \times [T_{pca}]_{d_p, d} \times [X_k]_{d,1}. \tag{3.10}$$

3.3.4 Classification

Nearest Neighbour (NN) [143] is the most commonly used classifier in gait recognition context as it is simple to perform effectively without need of training and Euclidean distance is used to measure distance between gallery and probe as below,

$$d_{ij} = \sqrt{\sum \left| F_i - F_j \right|^2}, \tag{3.11}$$

where, d_{ij} is the distance between i^{th} probe cycle and j^{th} gallery cycle and F_i, F_j are corresponding feature vectors.

To determine the similarity distance between two subjects, each of which is composed of multiple gait cycles, there are a number of approaches that have been used. One of the most common practices used in literature is to compute a single score for the similarity distance (SSD), where one optimum similarity score for a particular gallery subject with the probe subject is computed for performance metric. In this case, the distance to the closest gallery cycle from each probe cycle $\left(dmin_i^P\right)$, and the distance to the closest probe cycle from each gallery cycle $\left(dmin_j^G\right)$ is found,

$$dmin_i^P = \min_i \left\{ d_{ij} \right\}, \qquad dmin_j^G = \min_j \left\{ d_{ij} \right\}. \tag{3.12}$$

The final distance, $D(p, g)$, between the probe subject, p and gallery subject, g is,

$$D(p, g) = \frac{1}{2} \left(\text{median} \left\{ dmin^P \right\} + \text{median} \left\{ dmin^G \right\} \right). \tag{3.13}$$

Independent score for similarity distance (ISD) is another way of calculating score, where each probe cycle is treated as a separate entity for evaluation. Based on this, the final distance, $d(p_i, g)$, of i^{th} cycle of probe subject p to gallery subject, g is computed as,

$$d(p_i, g) = \min_j \left\{ d_{ij} \right\}. \tag{3.14}$$

Evaluations in this thesis use a single score for similarity distance as this is commonly used in literature to make a fair comparison. However, numerical values resulted using ISD scores are also reported in Appendix A.

3.3.5 Performance Metrics

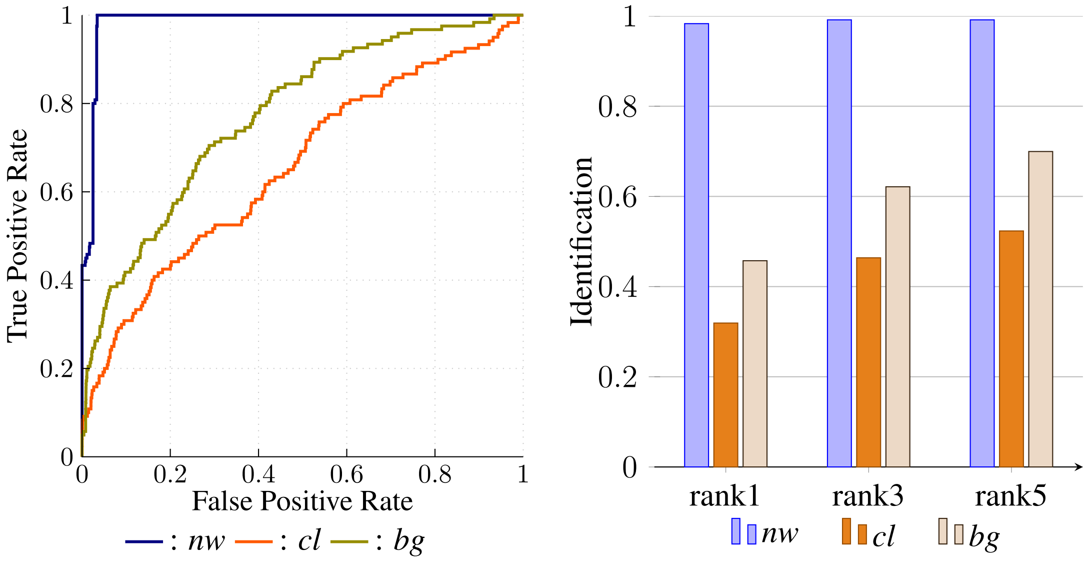
Computed similarity distances from the classifier are used to evaluate the algorithms for two different tasks, verification and identification, as explained in Section 2.2 in Chapter 2. To compare the identification performance of the algorithms in our research, cumulative match scores (CMS) are used. The rank in which the correct match of the particular probe within the gallery is found by computing the order of the score for the particular matched probe and the gallery. The percentage of the probe cycles matched with their pair gallery subject's gait cycle within the specific rank is used to compare the over-all performance. Cumulative match scores for Rank 1-10 are used in our evaluation based on the number of subjects in the database.

To compare the verification performance, receiver operating curves (ROC) are used that compare the true acceptance rate (TAR) with false acceptance rate (FAR). It illustrates the performance as a percentage of people who can be truly verified when a specific percentage of false identified people are allowed. Again true acceptance rates for false acceptance rate of 1-10% are used to evaluate our algorithms.

3.3.6 Experiments and Results

The proposed baseline is evaluated on the CASIA database as it provides multiple variants for inter-class cases with an adequate number of subjects. The performance matrices on other databases are reported in Appendix A.

GEI features are extracted and transformed to discriminative domain using PCA and MDA. Similarity distance scores are computed using an SSD approach. ROC curves and rank results based on these scores are shown in Figure 3.11. Identification rate of 100% at rank 1 and full AUC in roc curves are achieved for the intra-class case. However, this rates significantly drops in inter-class test cases as the features fails to handle appearance changes due to clothing and bag.



(a)

(b)

Figure 3.11: Recognition performance on the CASIA database. (a) ROC curves and (b) rank results.

To handle this issue, improved segmentation and dynamic static region analysis are explored on GEI features.

3.4 Enhancements to the Baseline for Robust Gait Recognition

Significant performance degradation has been observed on the GEI-based technique with the appearance changes from the evaluation on the CASIA database. This is because of its nature of encoding static and dynamic gait patterns. However, it shows the potentiality of recognising ability with its 100% rank 1 results on the intra-class case. In real world gait recognition systems, it is not always possible to get the same appearance gait features as they are enrolled. Therefore, to maintain the higher recognition performance in inter-class cases that can be influenced by clothing, carrying and segmentation noises, enhanced improvements that can be made to various elements in the implemented framework of the baseline are explored.

3.4.1 Influence of Improved Segmentation

Most of the gait recognition methods discussed in Chapter 2 require segmentation of silhouettes as their initial steps. Hence, accurate segmentation plays major role in gait recognition, particularly in an appearance-based method. However, most of the research in literature focuses on feature extraction and classifier steps and used distributed silhouettes with the database as it currently exists. The silhouettes distributed with the most common databases are resulted from simple background subtraction algorithms, such as frame differencing with thresholding used in [55]. Though such an implementation may be adequate for certain indoor scenes (all the three gait databases obtained are captured indoor), distributed silhouettes have high segmentation noises with cluttered background. When this issue is explored on the CMU MoBo dataset, significant shadowing, cluttered background with subjects' clothes sensor noise as well as compression artefacts, has been observed.

In the CASIA dataset, there appears to be an issue with the capturing system, where the image is dark and heavily biased towards the green. This problem is consistent though, and therefore does not impact (apart from the lowered contrast) the extracting of silhouettes. In a few select sequences however, for a few seconds, the video will revert to a more natural appearance, making background subtraction almost impossible.

Recently, graph-cut based segmentation algorithms with adaptive background modelling [84] show improvements over the existing algorithms and accurate cleaned foreground segmentation has been reported. Based on that, the method similar to [84] has been applied, with necessary modifications to obtain the accurate cleaned silhouettes. The remainder of this section explains underlining theory and implementation, and compares the new obtained silhouettes and assesses the performance on subsequent gait recognition.

Binary graph-based image segmentation

In graph cut based segmentation, by first, image is represented as a Markov random field (Markov Random Field (MRF)). The undirected graph, $G = (V, E)$, is constructed with a set V of vertices for each pixel in the image, and a set E of edges joining the vertices of neighbouring pixels.

Each pixel is also connected to two special vertices, called the source (s) and sink (t), which for this problem represents the foreground and background respectively. By performing a graph cut which separates the s and t vertices onto separate graphs (S

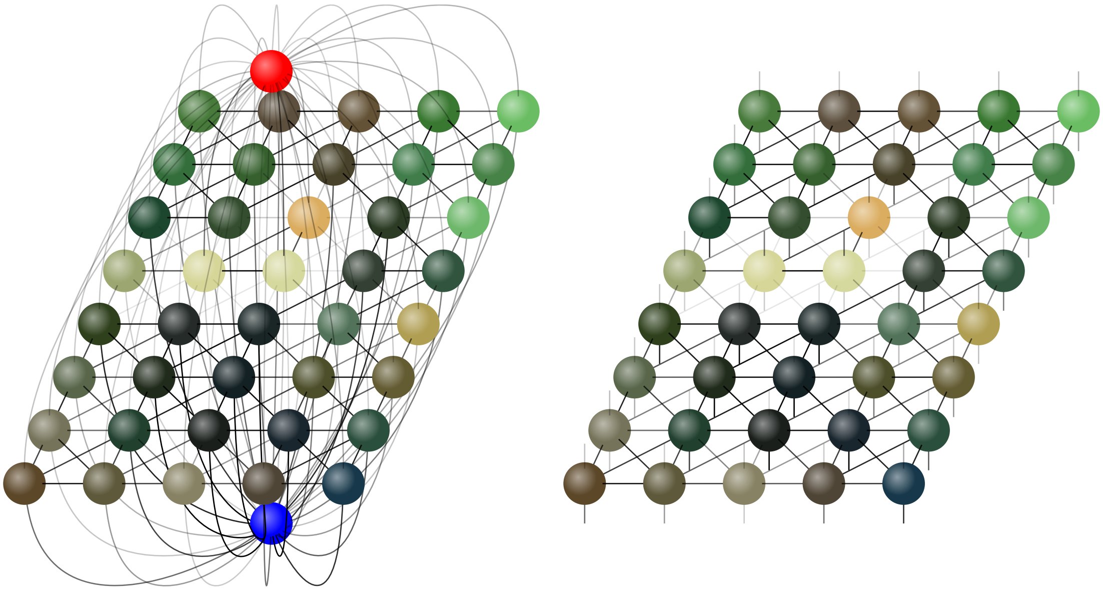


Figure 3.12: Image segmentation using graph cuts. (*left*) Graph structure of an example segmentation problem. Edge lines are weighted based on capacity values. (*right*) The same graph is shown but with the source and sink vertices removed for clarity.

and T respectively, such that $S \cup T = G$, $S \cap T = \varnothing$), segmentation can be achieved by determining which resulting graph on which the pixel lies, such that the output mask, M, at pixel, u, is represented by,

$$M\left(u\right) = \begin{cases} 1 & u \in S \\ 0 & u \in T. \end{cases} \tag{3.15}$$

An illustration of the graph structure of an example segmentation task is shown in Figure 3.12. A (heavily down-sampled) image is represented by the nodes, appropriately coloured. Each pixel can be seen forming edges with its eight neighbours, as well as the foreground (red) and background (blue) node.

Figure 3.13 shows the same problem after performing the graph cut. It can be seen that the original graph has been split into two graphs, with the source and sink nodes separated onto different ones.

This graph cut is typically performed by performing the minimum cut. The graph cut is driven by weights assigned to the edges, which can be seen as a cost value incurred, should that edge be removed. The cut is made to the graph such that the energy, or total cost incurred, of the cut is minimised. The algorithm finds the globally optimal solution, achieving a ‘minimum cut’.

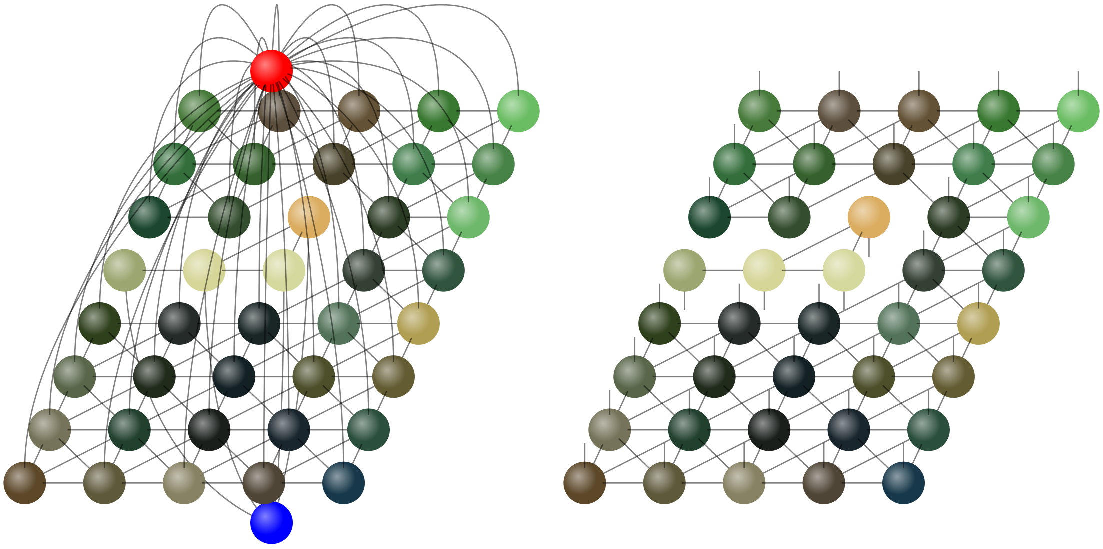


Figure 3.13: Graph cut via mincut. The resulting graphs after graph cut (*left*) with and (*right*) without the source and sink vertices.

The edge weights are derived from the joint probabilities between vertices in the MRF, which is the likelihood of the two vertices occurring together in the same region, *i.e*. the capacity value, c_{uv}, of the edge uv is determined by,

$$c_{uv} \propto \mathrm{P}(u, v \in X), \tag{3.16}$$

where X can either be the foreground (FG) or background (BG) region. For the weights between neighbouring pixels, this value is usually derived based on similarity in pixel appearance, typically simply the distance between the pixels' colour values. Pixels with a close similarity are more likely to belong together in the same region, and thus given a higher cost value, reducing the potential they will be separated in the graph cut.

Similarly, the foreground and background weights (c_{ut} and c_{us}) are determined by the pixels' likelihood in belonging in the foreground and background. As the s and t vertices can only exist in their own region, equation (3.16) can be expressed as,

$$\begin{aligned} c_{us} &\propto \mathrm{P}(u \in FG) \\ c_{ut} &\propto \mathrm{P}(u \in BG). \end{aligned} \tag{3.17}$$

These edge weights are computed based on global colour model by finding the similarity between the edge pixels' colour values. In Figure 3.12, the weights are represented by the density of the edge, where a higher weight value results in a darker line.

The consequence of using the minimum cut to perform the segmentation results in a preference for silhouette boundaries to occur along the edges of distinct regions, helping to group pixels of similar appearance together. This is because the joint probabilities of the pixels across the boundary are generally lower. This method of segmentation is likely to result in cleaner silhouettes versus one where the foreground and background probabilities are simply compared, particularly in the cases where the models themselves are not very accurate.

Graph cut is used for background subtraction as follows; The Euclidean distance in pixel values between the image and the background is used to weight graph edges; the larger distance results in a higher foreground weight. For the background weight, threshold value, T, is used. For simplicity, the distance value is directly used, such that the graph capacity values become,

$$\begin{aligned} c_{us} &= T, \\ c_{ut} &= \left| I\left(u\right) - B\left(u\right) \right|, \end{aligned} \tag{3.18}$$

where a colour channel range is assumed as 0–1.

Colour correction by histogram matching

Even in controlled set-ups such as in the datasets mentioned in Section 3.2, background miss-match can exist due to changes in camera exposure and applied gain. However, camera-auto-mode adjusts parameters depending on total light captured. Moving objects also can change overall brightness of the scene. To correct these variations, histogram matching is used as a pre-processing step, in order to transform an image such that the background appearances are normalised.

Given the controlled nature of the experiment here, the background can be guaranteed to remain static, and since any apparent changes to its pixel values can be attributed to an effective applied gain (or noise), normalising the colour distribution between the images should be able to correct for any differences in the background.

To perform the transform, histograms of the background for both the target and source images need to be calculated. Given that the goal is to segment away the background, this may seem premature, though for the purposes here, only a part of the background is needed, with a few caveats. Firstly, the pixel locations used must be the same between the two images. Care needs to be taken to select only background

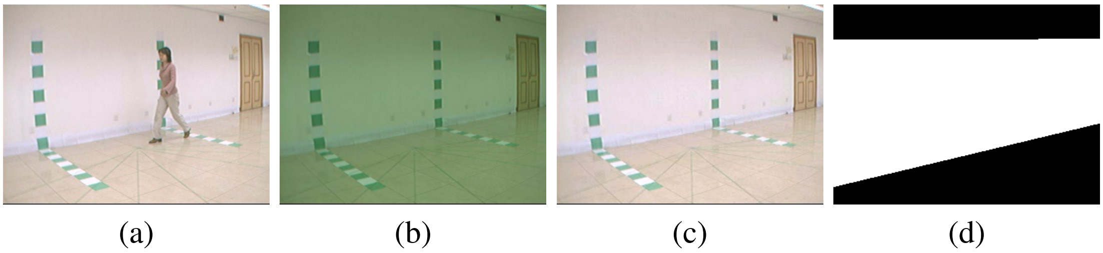


Figure 3.14: Example colour distorted images. (*a*) Foreground and (*b*) background image. Colour discrepancy between the two makes it difficult for accurate segmentation to be performed. (*c*) Colour corrected background image using histogram matching, with the (*d*) region mask used for this dataset.

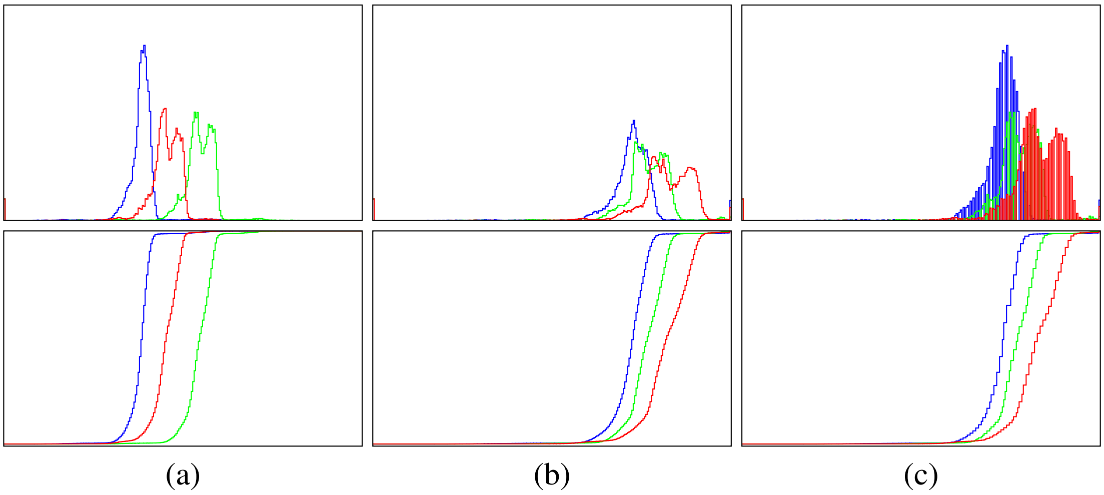


Figure 3.15: Image colour histograms. Histograms (*top*) and cumulative histograms (*bottom*) of the (*a*) source and (*b*) target images, as well as the (*c*) corrected image after transformation. The differences in the corrected image's histogram are due to the discrete colour values.

pixels, and only those that lighting effects, such as shadows are consistent between the two. Finally, sufficient representation of the colour profile is needed; not necessarily that the actual distribution is obtained, but that enough samples have been selected at each colour value so an accurate mapping can be determined.

An extreme example from the CASIA dataset is shown in Figure 3.14, where motion image is dark and heavily biased towards the green and background image appears as more natural. Since the motion path of the foreground subject is predetermined, a single, manually labelled mask has been used throughout the entire dataset for each view. Following the previously stated conditions, the mask masks out the region where motion is expected to be, leaving behind the static background pixels.

Histograms (Figure 3.15) are constructed of the background pixels for both the

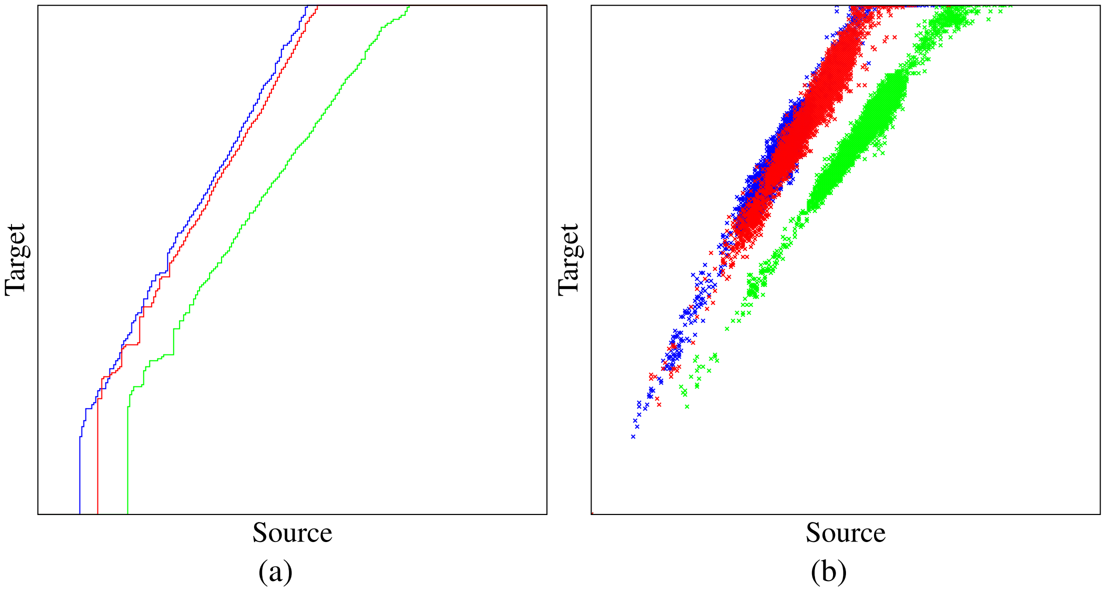


Figure 3.16: Colour map resulted from colour-histogram matching. (*a*) Calculated intensity transform for all 3 colour channels from histogram matching. (*b*) Background pixel intensity values of the source image plotted against the target.

source and target images, with each channel treated independently. A transform is created that maps the pixel intensity values of the source image to the target. The mapping is calculated, such that the cumulative histogram of the transformed image matches that of the target. The resulting map is shown in Figure 3.16. This map is applied to the entire source image, with the resulting image seen in Figure 3.14.

A more common scenario is where the difference is more subtle and is not clearly noticeable at first, but upon closer inspection, small differences in colour can be seen in the background, as well as changes in the brightness of the treadmill. The background image is corrected to match the colour profile of the foreground image. Silhouettes are extracted using background subtraction with a low threshold (0.05) to highlight the differences.

Results

Example segmented silhouette outputs from the CMU MoBo database are shown in Figure 3.17. Thin treadmill outlines can be seen in some of the silhouettes. This is a result of the treadmill flexing slightly due to the applied load. This is the result of a slight downward displacement of the treadmill due to the applied load. Overall, the overwhelming majority of errors are due to false negatives, where the foreground regions have been misclassified as background due to the difference in colours between

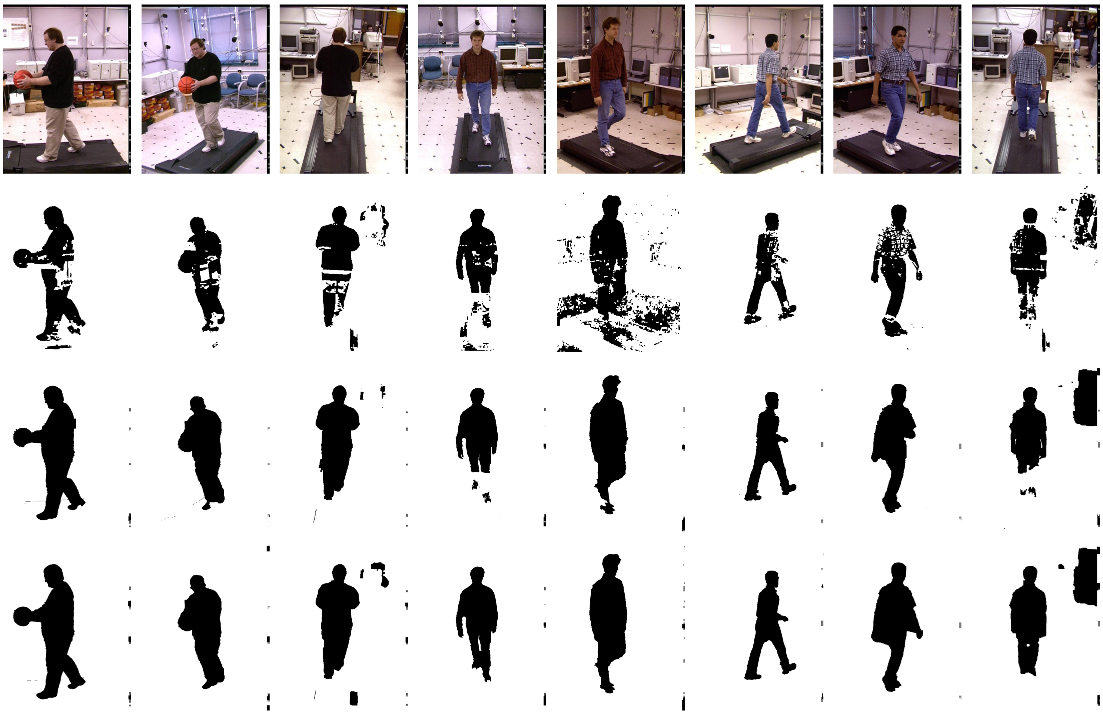


Figure 3.17: Segmentation examples on the CMU MoBo database (*top* to *bottom*) RGB images, Thresholding, graph cut without and with colour correction.

them to be lower than the threshold value.

The optimal threshold values used here appear quite high and are not lower likely due to the increased errors from false positives not being able to overcome any potential benefits. This suggests a likely disparity between the test and background images, either due to noise, or some other factor. After all, the threshold used here for background subtraction is simply a tolerance value to compensate for any noise. Under a perfect, noiseless case, the optimal threshold would be 0, since background regions would remain identical, while any difference in pixel value would indicate a foreground region.

An example is in the 5^{th} image of Figure 3.17, where significant errors in the background occur for little apparent reason. A visual inspection of the images shows that there is a brightness and contrast imbalance between the image and its corresponding background. This issue occurs to some degree in nearly all images, but is most prevalent for that camera view. However, a reduction in errors can be seen when using the corrected background using colour correction technique as it lowers the offset imposed due to the differences inherent in the image.

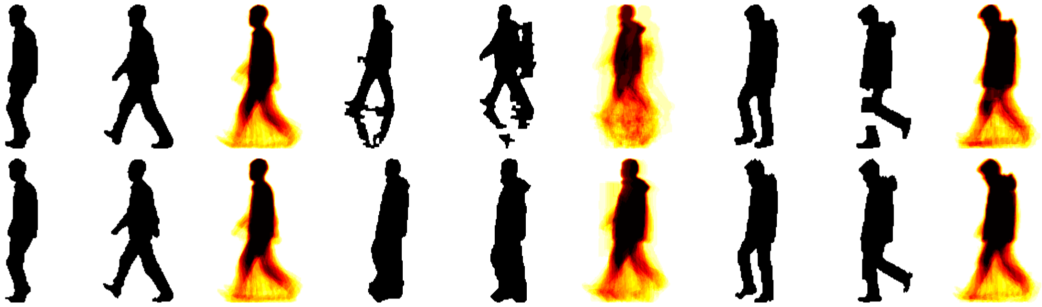


Figure 3.18: Comparison of silhouette quality on the CASIA dataset. Silhouettes and resulting GEIs using the (*top*) original silhouettes in the dataset and (*bottom*) segmentation algorithm presented in Section 3.4.1.

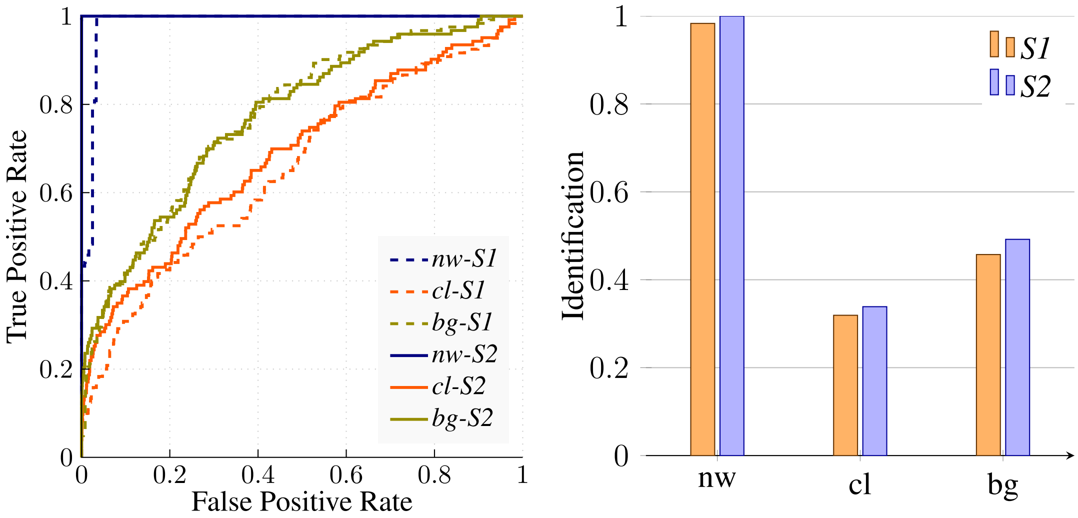
Influence of segmentation on gait recognition

Figure 3.18 compares the output GEIs from the original silhouettes distributed with the CASIA database and the cleaned silhouettes resulted from the graph-cut (Section 3.4.1). It is noticed that segmentation issues, particularly due to the shadows and colour similarity, make frame registration fail and are heavily reflected in corresponding GEIs.

To evaluate the influence of segmentation on gait recognition performances, experiments are conducted on the CASIA database. ROC curves are shown in Figure 3.19a and rank-1 identification results are shown in a bar plot 3.19b. Consistent improvement of $5-10\%$ in recognition rates using cleaned silhouettes ensure that recognition performance using GEI is purely from gait appearance rather than unique noise attached to an individual; also, poor segmentation can degrade the GEI-based gait recognition system. This confirms the potentiality of the GEI feature in distinguishing an individual. However, inter-class cases still perform comparatively poorly. Further analyses are carried out to identify the GEI regions that are more sensitive to external appearance changes and the regions which contribute to distinguish the individuals.

3.4.2 Dynamic Region Analysis on GEI

The potential problem of appearance-based method particularly on GEI is that changes in appearance make major impact in extracted features. Changes in personal condition such as carrying bag, wearing different types of cloths make significant appearance changes particularly to the upper body. To explore their influence on the walking style of an individual, it is vital to remove the appearance changes made by them. However, removing them will also reduce gait features as they are occluded with human body.



(a) (b)

Figure 3.19: Comparison of recognition performance using different segmented silhouettes on the CASIA database. Recognition performances using the silhouettes resulted from the proposed segmentation algorithms (*S1*) and the silhouettes distributed with the database (*S2*) ((a) ROC curves and (b) rank results). *nw* is the intra-class test case using ‘normal walk’ and, *cl* and *bg* are inter-class test cases, where the probe subjects with ’different clothing’ and ’carrying bag’ are compared against the gallery, ‘normal walk’, respectively.

Recently, many techniques have been introduced as an extension of GEI to handle these appearance changes. However most of them are concerned with a particular type of condition at a time. To make a generalised solution, the influence of the dynamic and static nature of GEI on gait recognition performance relates to the appearance changes are investigated first. Then, the influence of removing GEI regions where external appearance mostly occurs is explored.

Separation of dynamic and static regions

Recognition performance of appearance-based techniques, particularly GEI, is based upon the implicit notion of what is being observed [48]. It represents temporal movements as well as spatio positions and shapes of human body parts. That is a combination of static and dynamic parts of gait. Though both contribute to discriminate individuals, static pixels are more sensitive to external appearance changes such as changes in clothing, carrying bags *etc*. Therefore balancing between static and dynamic representation of GEI gait recognition is essential, particularly in inter-class matching.

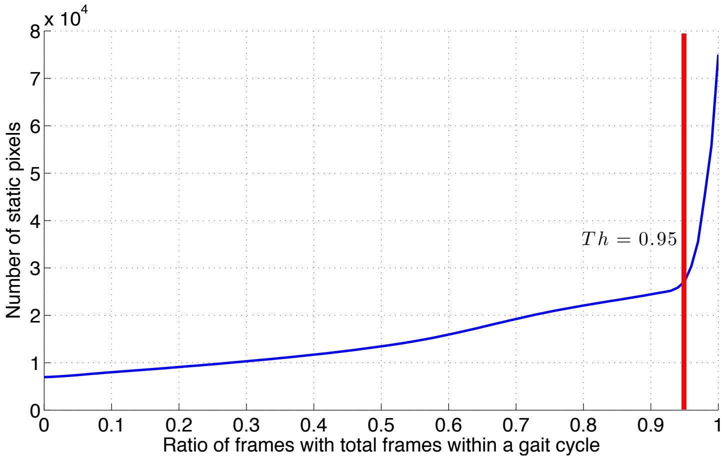


Figure 3.20: Averaged pixel distribution with different re-occurred frame ratio (Th)

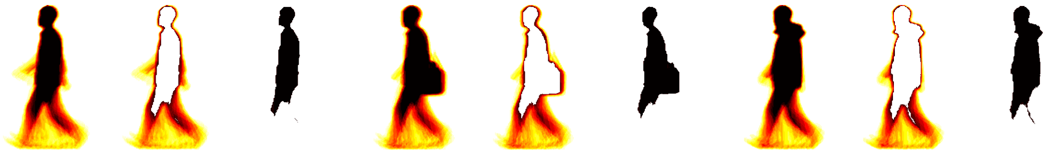


Figure 3.21: Example GEI and separation of dynamic and static portions (The three gait challenging conditions (*nw*, *bg* and *cl*) in the CASIA database are considered).

Static and dynamic parts are distinguished based on how they change with the time. A part that re-occurs in the same spatio-phase on the aligned silhouette within a gait cycle is defined as static (Ψ) and others are dynamic (Δ). However, there are chances of missing bits, segmentation noises and improper registration in silhouettes. To tolerate this, re-occurrence of pixels is only counted for Th ratio of the number of frames in a gait cycle as follows,

$$\begin{aligned}\Psi_{GEI} =& GEI(GEI \geq Th) \\ \Delta_{GEI} =& GEI(GEI < Th)\end{aligned} \tag{3.19}$$

To define the correct Th value, the distribution of static pixels are analysed with the different ratio of re-occurrence. Figure 3.20 plots the distribution pixels with a ratio of occurrence that is averaged for training GEIs from the CASIA database view $90°$.

Based on the plot, it can be noted that pixels that reoccurred in more than 95% of total frames within a gait cycle makes significantly higher than with the other variations in Th. Therefore $Th = .95$ is chosen for approximate threshold value to separate static and dynamic pixels. Example GEI, Δ_{GEI} and Ψ_{GEI} are shown in Figure 3.21.

Contribution of dynamic and static regions on gait recognition

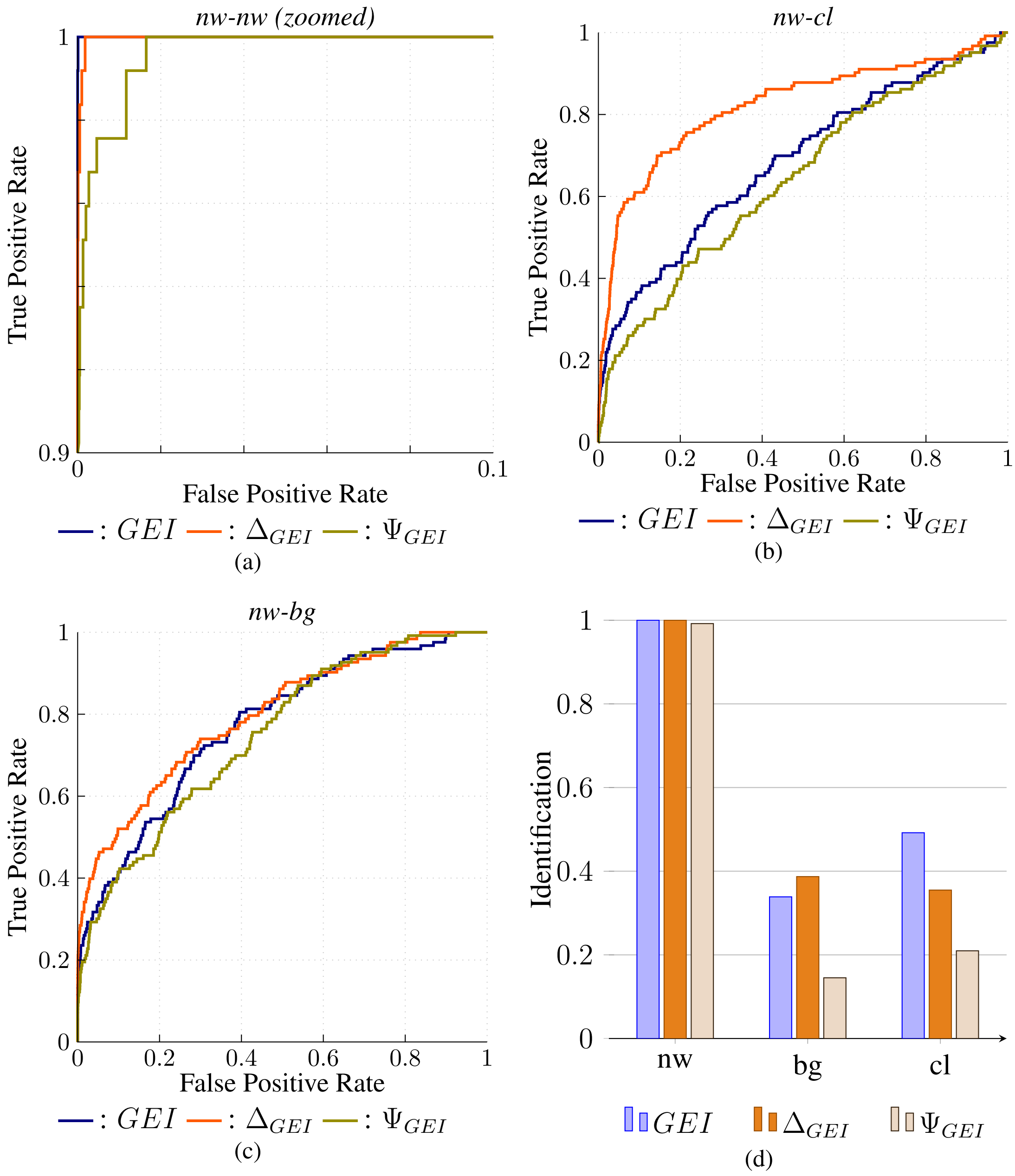
The contribution of both static and dynamic regions of GEI on gait recognition performance are explored separately, to understand the discriminating nature of each region. Figure 3.22 compares ROC curves and rank-1 recognition performance, using these separated dynamic and static GEIs with the original full GEI. From the curves, it can be noticed that removing static parts doesn't degrade recognition performance, rather, this improves it. A $15 - 20\%$ improvement in verification rate in inter-class test cases, using only the dynamic pixels, showed that static parts of GEI are mostly affected by the appearance changes, due to clothing and carrying things. In a clothing case, it significantly boosts the verification rate, as static noises caused by clothes are removed in Δ_{GEI}. However, removing the static region reduces identification rates in bg condition, as an oscillating bag affects dynamic representation of the gait and that is more emphasised when only dynamic parts are used.

3.4.3 Analysis of Spatial Contribution of the GEI

Based on the static and dynamic region analysis, it has been observed that Δ_{GEI} alone can better represent gait. Also it has been noticed that this dynamic region mostly occupies the lower part of the silhouette. Based on the inter-class differences in available gait databases, appearance changes, such as clothing and carrying things, affect mostly the upper body. To solve the issue of dynamic changes of appearance noises on Δ_{GEI}, the above three observations are used to explore the influence of GEI regions based on their height.

To do the height-based analysis, GEI is segmented with the patch size of 5 throughout its height as shown in Figure 3.23(a), and dynamic-to-static pixel ratios (R_{Δ}) for each segment are computed. Figure 3.23 shows the variation of this ratio with respect to height from the bottom for an example subject on the CASIA database with different classes. Within the lower 40% of silhouette, dynamic-to-static ratio changes smoothly and consistently with the external appearance changes. However, this varies significantly in upper region static ratio dominates ($\Psi > \Delta$).

Based on this analysis, the GEI region where R_{Δ} exceeds R_{Ψ} is selected as better gait feature that is less susceptible to appearance changes. During the analysis on the CASIA database this region approximately occurs on 40% of the height of the silhouette from the bottom. Figure 3.24 plots height where R_{Δ} exceeds R_{Ψ} for the example GEIs computed on the CASIA database clothing condition. Computed h is



(a)

(b)

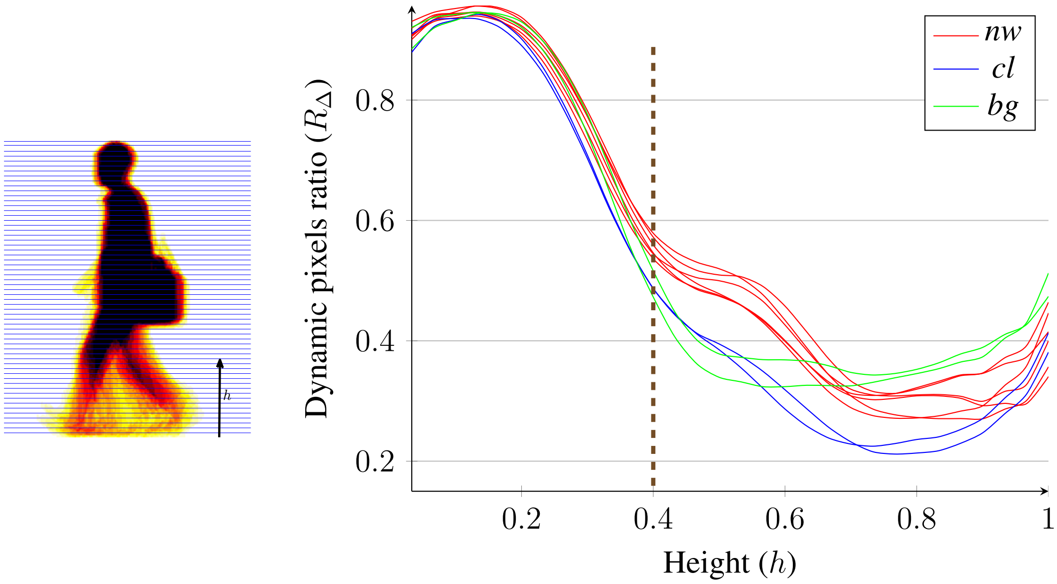
(c)

(d)

Figure 3.22: Comparison of dynamic and static nature on GEI on the CASIA database ((a),(b) and (c) are ROC curves and (d) identification rates at rank 1).

Type	$\mu(h)$	$\sigma(h)$
nw	0.4138	0.0145
cl	0.4022	0.0393
bg	0.3812	0.0421

Table 3.3: Comparison of height (h) values computed for cropping the upper body.



(a) (b)

Figure 3.23: Variation of dynamic pixel ratio, R_Δ with height h.

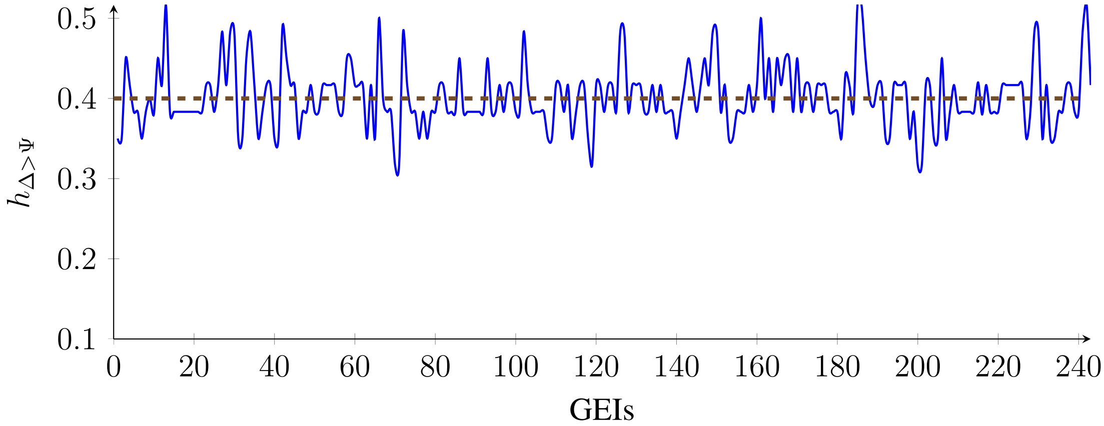


Figure 3.24: Variation of average $h_{(\Delta>\Psi)}$ for the silhouettes on the CASIA database.

consistent with mean value around 40% of height with ignorable variance in all the test cases as shown in Table 3.3. The GEI that is truncated to the height h is named as truncated GEI, $TGEI$, and has the following potential merits over the traditional GEI:

- Computational efficiency increases due to the dimensionality reduction (Dimension is reduced by as more than two times).
- Contribution of static pixels that corresponds to legs is still included.
- Less susceptible to the most common appearance changes.

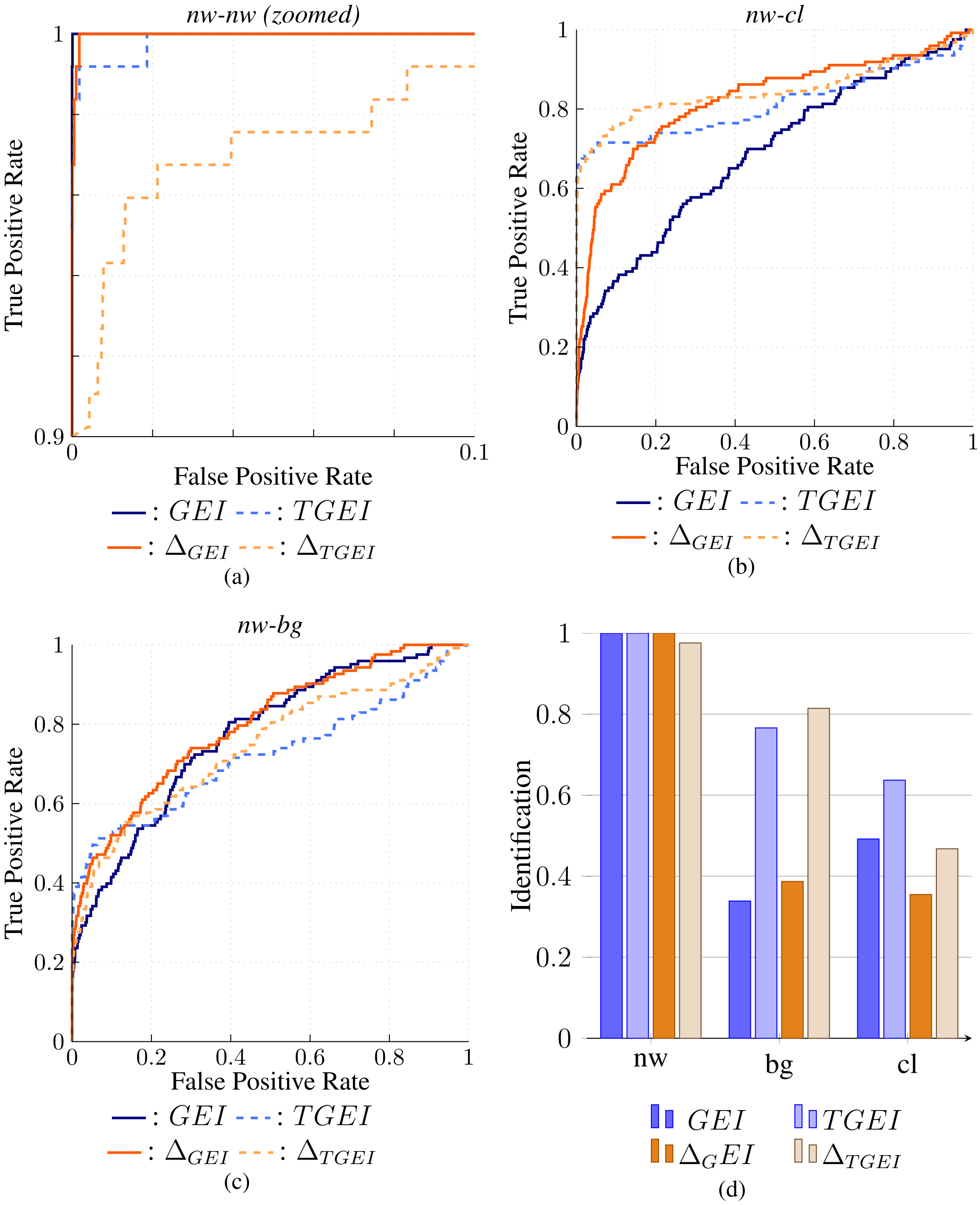
Experiments on truncated GEI

Experiments are carried out on the CASIA database to evaluate the proposed truncated GEI. Height to generate the truncated GEI is computed, based on the dynamic behaviour of the training data. Figure 3.25 shows ROC curves and rank1 results on the CASIA database.

3.5 Summary

This chapter outlines how the baseline works to implement the elements in the gait recognition framework and illustrates further modifications and extensions to improve the gait recognition performance. GEI is chosen as a feature extraction baseline, because of its simplicity and improved recognition performance, as reported in recent literature. At first, the GEI-based baseline algorithm is implemented and its limitations are assessed. As GEI is more susceptible to appearance changes, enhanced modifications are proposed to increase the robustness of the algorithm.

With the analysis on the segmentation algorithms, consistent improvement of $5-10\%$ in recognition rates using cleaned silhouettes confirms that the GEI encodes better discriminating information of an individual, and robust segmentation is needed to perform appearance-based gait recognition. It has also been found that truncated GEI with the height of 40% of the GEI shows equal or greater gait recognition performance while handling the gait challenging conditions. With these preliminary analyses on the baseline implementation, 50% improved results in the test case of *bg* and 20% higher recognition rate in the test case of *cl* on the CASIA database are achieved, compared to the performance reported in [98].



(a)

(b)

(c)

(d)

Figure 3.25: Comparison of GEI and Δ_{GEI} on truncated height using the CASIA database ((a),(b) and (c) are ROC curves and (d) identification rates at rank 1).

Chapter 4

3D Gait Recognition using Multi-view Data

4.1 Introduction

View dependency is a major issue in 2D gait recognition techniques. In 2D, gait dynamics are presented mostly in side-view, since self-occlusion between legs is minimised. Due to that, particularly in 2D appearance-based techniques, performances heavily degrade when the view angle or walking direction of the subject changes from side-view. One of the main reasons for this, is that image frames or extracted features cannot be aligned with the relevant parts properly and feature correspondents vary a lot due, to the self-occlusion on images. That makes it impossible to match them to perform classification.

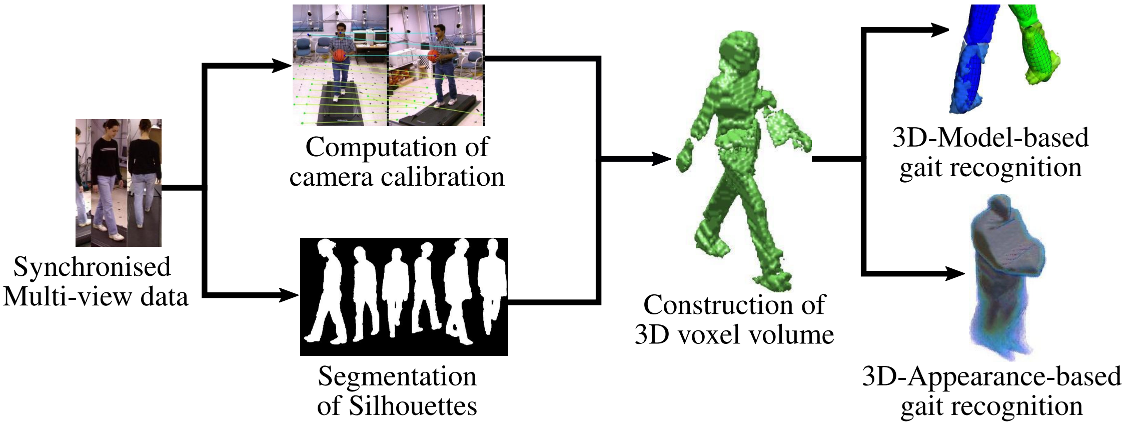


Figure 4.1: Gait analysis on multi-view data

One of the possible solutions to enable recognition of subject with different walking directions is reconstruction of a 3D model of the subject. The move to 3D space solves the issue of view dependency and problems with self-occlusions, though this method does constrain its applications to situations where a multi-camera set-up is in existence.

During the research work, model-based & appearance-based approaches in 3D are explored, since both of them have their own advantages. Figure 4.1 illustrates the overall view of gait analysis on multi-view data explored in this chapter. The initial phase of working in 3D for gait recognition is reconstruction of a 3D model of the subject using the synchronised multi-view data. Two multi-view databases, CMU MoBo database and CASIA database are used to explore gait in a 3D context. In order to reconstruct the voxel-volume from these multi-view databases, camera calibration details are computed. Using the reconstructed 3D voxel-volume, both appearance-based and model-based approaches in gait context have been explored.

4.2 3D reconstruction from multi-view images

A 3D reconstruction of a subject is done by constructing voxel volumes and defining each voxel by checking the existence of its projection on 2D binary silhouettes in each view.

To do this, at first, cleaned binary silhouettes are extracted from corresponding multi-view images using optimised graph-cut based segmentation, as explained in Section 3.4.1 in Chapter 3. Examples of multi-view segmented silhouettes from the CMU MoBo database are shown in Figure 4.2.

A 3D reconstruction of human silhouettes required camera calibration details to project 2D coordinates in different views to 3D. The remainder of this section explains the computation of camera calibration and volume reconstruction.

4.2.1 Computation of Camera Calibration Details

Camera calibration comprises intrinsic (internal camera parameters such as focal length, image centre, camera centre) and extrinsic (rotation and translation parameters) parameters of a camera used to capture the images. If camera parameters are not known for time synchronised images, camera calibration matrices can be produced by analysing the detected point correspondences manually or automatically.

Both the available multi-view gait databases (CMU MoBo [31] and CASIA [55] are not provided with calibration details. According to the literature review, there is

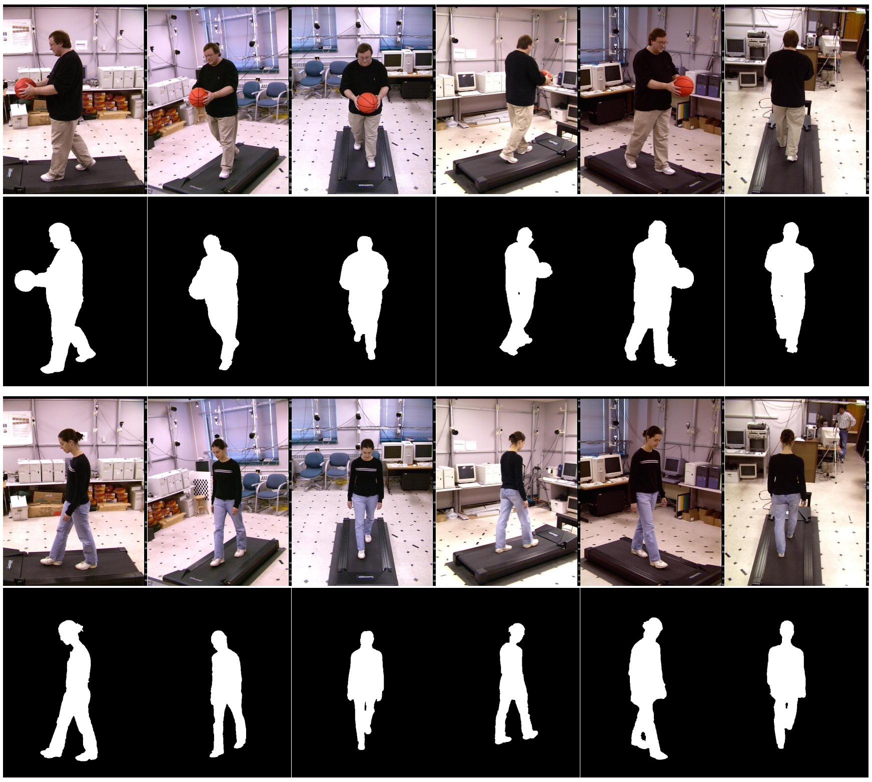
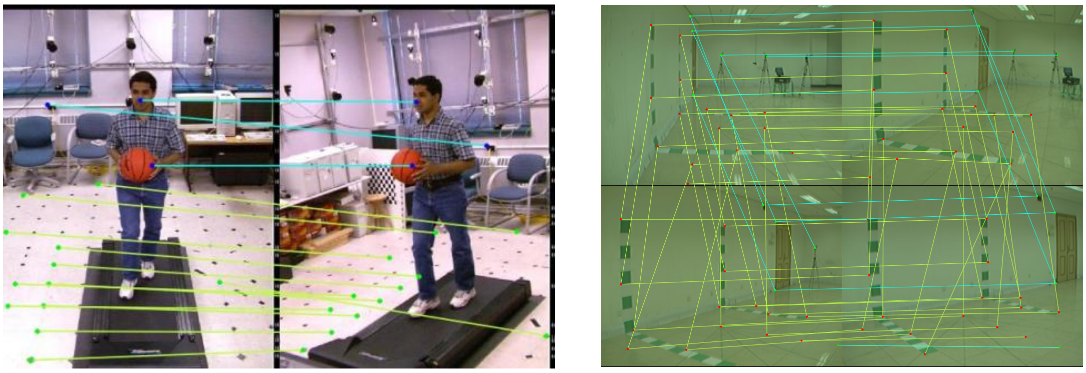


Figure 4.2: Motion detection and background subtraction. Examples resulted using graph cut [83].

no common multi-view data suitable for 3D reconstructed models because of lacking calibration details. A point correspondences approach for computing the initial estimation of calibration was applied and fine-tuned using particle filters. Initially, point correspondents with known and unknown 3D coordinates were manually selected as shown in Figure 4.3. From the known 3D coordinates, initial calibration matrices are produced. Then, using these calibration matrices, 3D coordinates for all point correspondents are computed. Again these 3D coordinates are projected back to 2D and deviation between actual and projected coordinates are computed to define the weight for the particle filters. Using these weighted particle filters, again the new calibration matrices are computed and this process is iteratively repeated to minimise the variance between actual and projected coordinates. Computed calibration details are listed in Appendix B to make the usability of the multi-view data of the CMU MoBo and the CASIA databases for future research.



(a) CMU MoBo Database.

(b) CASIA Database.

Figure 4.3: Camera Calibration using point correspondence (— : 3D known coordinates, — : 3D unknown coordinates.)

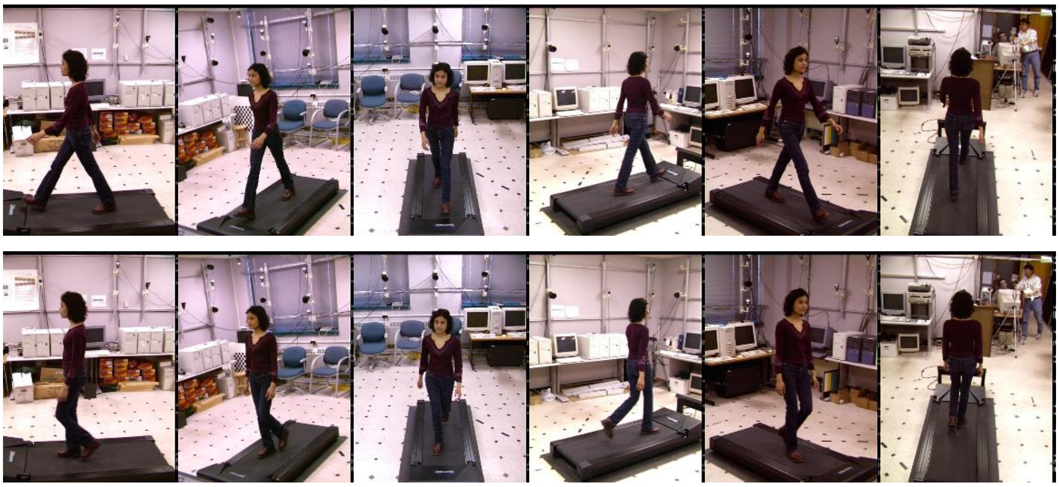


Figure 4.4: Examples of synchronised multi view images from the CMU MoBo database [31].

4.2.2 3D Reconstruction of Voxel Volume

To compute the 3D reconstruction from multi-view data, it is vital to have accurate calibration details with the time-synced data from each view. Images in different views of CMU MoBo database are synced when it's captured, as shown in Figure 4.4.

Though the CASIA database is captured in 11 views, each view is aligned with a different timing index as shown in Figure 4.7(a). However, effort has been made to correct the sync issues between views on the CASIA database since it is the only multi-view data with a larger number of image sequences. As an initial step, offset frames with respect to 90^o view are computed using a semi-automatic syncing process, using dynamic features in the frames and movement of 3D transformation of head. Head

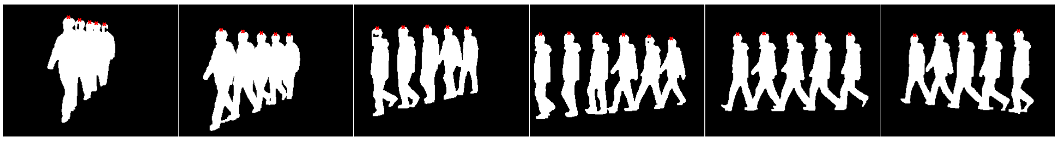


Figure 4.5: Examples of head tracking in 2D.

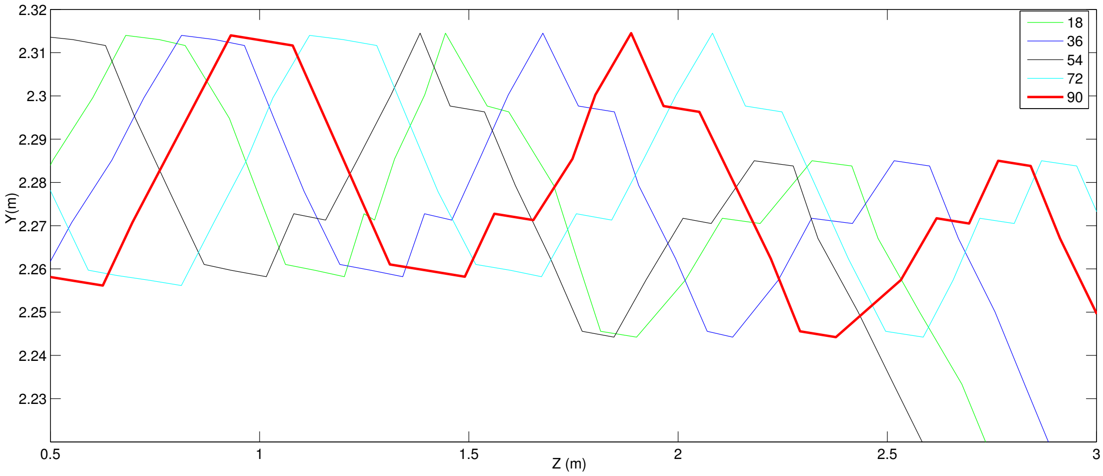


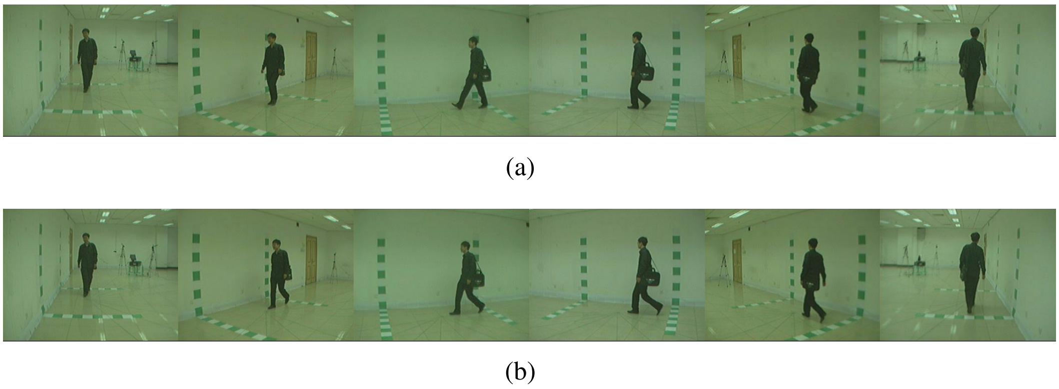
Figure 4.6: 3D head potions with respect to 90^o view.

positions are tracked on 2D image by finding the top position on extracted silhouettes as shown in Figure 4.5 These head 2D coordinates are projected to 3D line using the camera calibration matrices found in Section 4.2.1. Then, using the assumption that a person walks in a straight line with the fixed distance of $1.125m$ from the wall, a head position in 3D is computed. These coordinates are cross-correlated with respect to 90^o view and frame lagging is found as shown in Figure 4.6. Resulted frame offsets with 90^o view and the aligned images are shown in Figure 4.7(b).

The starting-offset alignment done on the CASIA database doesn't help in the proper syncing of multi-view images. Severe and random frame skipping in the middle is observed, particularly in $108 - 180$ views as shown in Figure 4.8.

Due to the random frame skipping in the CASIA database, algorithms in the remaining of this chapter are explored only on the CMU MoBo database.

The 3D reconstruction of the full human silhouette can be done by projecting pre-processed silhouettes and forming a visual hull, from which a rough voxel model of the subject can be constructed. The axis domains are referenced as, vertical to z-axis, while the x-axis represents the direction of motion. Using intrinsic and extrinsic camera calibration, points in the image plane and a real-world coordinate system are



(a)

(b)

Figure 4.7: Example results of semi-automatic frame lag-offset alignment. Examples images at 50^{th} frame from views 0^o, 36^o, 72^o, 108^o, 144^o, 180^o (left to right) on CASIS database. (a): miss-aligned frames, (b) after corrected with the frame offsets of $2, 6, -1, 4, 4, 0$ with respect to 90^o view, computed using the proposed alignment algorithm.

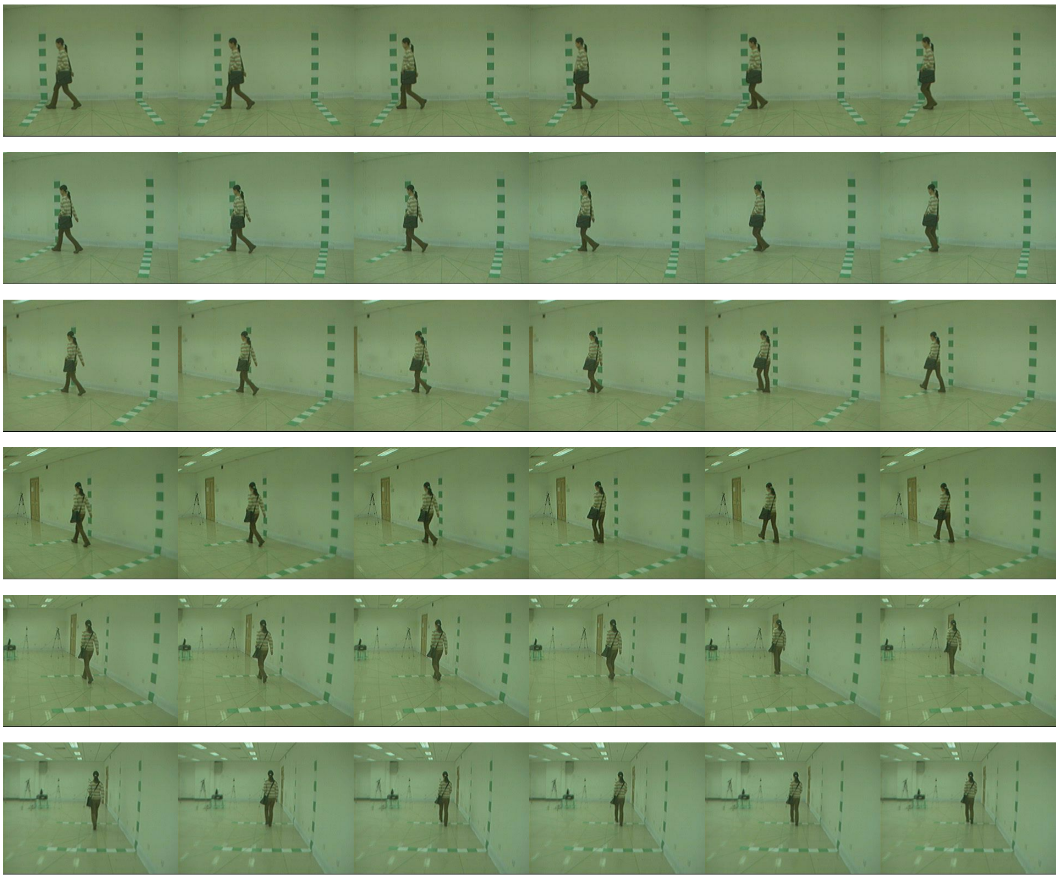


Figure 4.8: Frame skipping of multi-view images on the CASIA database. Each row corresponds to different view and column corresponds to frame index. Corresponding frames in the middle columns show different phase of walking in different views.

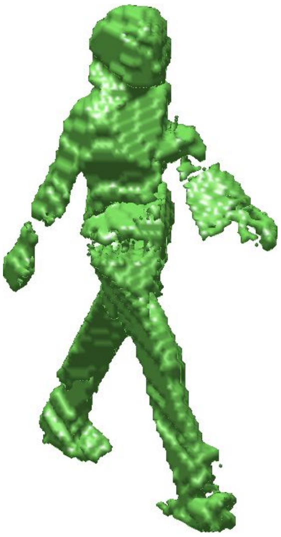


Figure 4.9: Example of reconstructed 3D voxel volume.

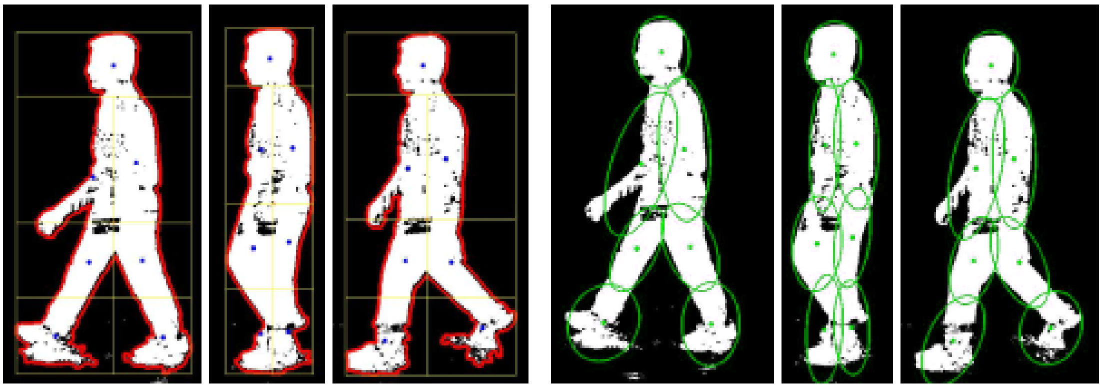
related as follow,

$$s \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_x & r_{xy} & r_x z & t_x \\ r_y x & r_y & r_y z & t_y \\ r_z x & r_z y & r_z & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4.1}$$

where (X, Y, Z) are the coordinates of a 3D point in a real-world coordinate space, (u, v) are the coordinates of the projection point in pixels. Intrinsic camera parameters are given by the principal point (c_x, x_y), f_x and f_y are the focal lengths and s is some scalar. To construct the 3D human silhouette, a 3D binary volume that can encapsulate the entire walking environment is built. Each point in the volume is projected into each view to check the corresponding 2D binary silhouette at the projected point. A reconstructed 3D voxel volume of human silhouette is shown in 4.9.

4.3 3D Model-based Gait Recognition

Model-based techniques gather the gait features by interpreting the gait dynamics in skeleton form, rather than indirectly, depending on the video data as in the appearance-based case. However, these approaches have difficult in fitting models accurately when the data is available in low resolution with segmentation noise [144]. In this case, ellip-



(a) Segmentation of Silhouettes.

(b) Ellipse fitting to the segmented silhouette.

Figure 4.10: Silhouette segmentation and ellipse fitting.

tical model-based approaches perform better because the simple enriched features are not very sensitive to these noises. Lee and Grimson [104] proposed ellipse fitting to the silhouette pixels and extracted moment-based region features. Rather than taking the entire silhouettes, they divided the silhouette into regions and fitted ellipses based on the statistics on the region. Because of the simplicity, good results and expandability, this ellipse fitting approach was chosen as the baseline algorithm. A novel ellipsoidal, 3D-based, view invariant algorithm is proposed that shows better performance.

4.3.1 2D Ellipse Fitting Baseline Implementation

Initially, binary silhouettes are extracted from the video sequences and pre-processed as mentioned in Section 3.4.1 of Chapter 3. The Region Of Interest (ROI) of the pre-processed binary image is computed and segmented horizontally and vertically with respect to the centroid. Then, the lower part of the ROI is evenly divided to separate the lower limbs and thighs. The head and torso in the upper part of the ROI are separated in the proportion of 1:2. All together there are seven regions (left and right parts of thighs, lower limbs and torso and head) defined, as illustrated in Figure 4.10(a).

Covariance matrices for each region are computed. These covariance matrices are decomposed into eigenvector and eigenvalue, which are used to define orientation, length and diameter of the fitted ellipse for the particular region, as show in Figure 4.10(b). From the fitted elliptical parameters, the following feature vectors are chosen.

- Ratio of major minor axis length/elongation (l).
- Orientation angle of major axis (α).
- Position of Centroid (x_c, y_c).
- Width of the bounding box (w).

4.3.2 Proposed Semi-dynamic 3D Ellipsoid fitting

A novel 3D based ellipsoidal fitting method with semi-dynamic segmentation is proposed that addresses the following issues in the elliptical baseline.

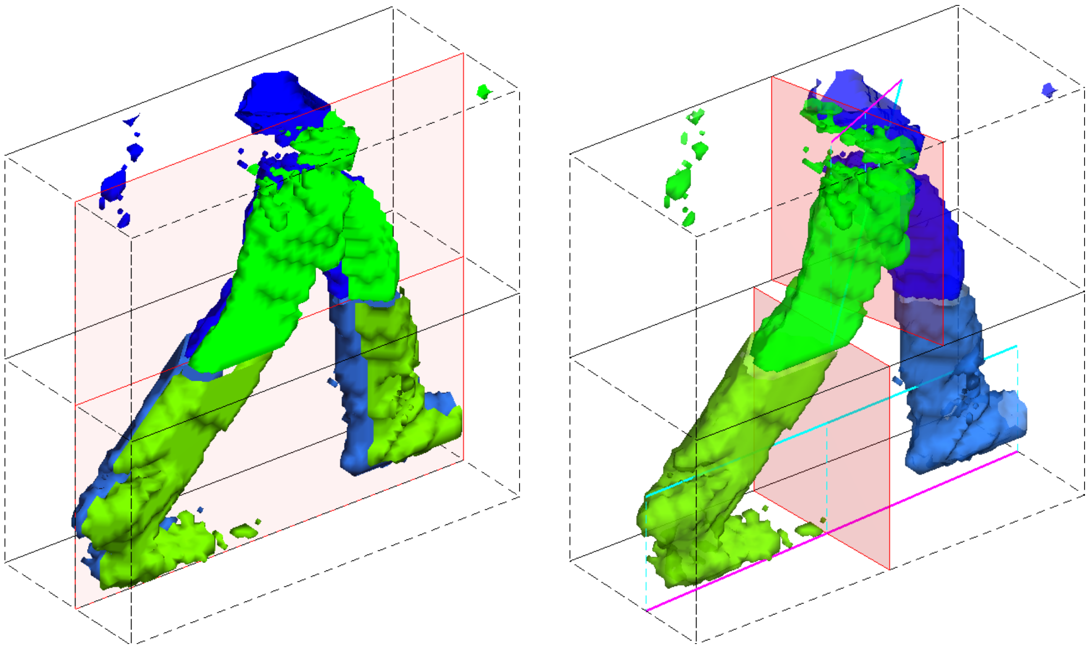
- The elliptical baseline can only work under orthogonal view.
- Traditional anatomical based segmentation fails to represent the actual gait pattern.
- Segmentation issues such as occluded hand and legs significantly affect on the recognition performance of the 2D baseline.

As an initial step, 3D voxel-volume of the subject is constructed as described in Section 4.2. The centroid of the 3D voxel volume is used to separate the upper and lower body as it provides a sufficient approximation of the hip/waist. To differentiate between the thigh and lower leg, the knee is approximated to be half way between the centroid and the bottom of the model.

In the 2D case presented in [104], the left and right leg cannot be differentiated when they occlude each other. With the 3D voxel volume presented here, this is now possible. However, within a gait cycle, the legs are not constrained to move along a single plane and can pass in front (and behind) each other. As a result, the left and right limbs cannot be separated simply along the xz plane.

In the proposed algorithm, the left/right differentiation of the upper and lower leg is performed separately. The algorithm described in Section 4.3.1 is used to find the major axis of the upper and lower volume distribution. The segmentation plane is chosen to be orthogonal to the projection of this major axis onto the xy plane (Figure 4.11).

Following the segmentation mentioned above, four segmented regions are selected, comprising of the upper and lower legs. Extracting ellipsoidal parameters for each region involves computing the mean and covariance of the voxel distribution in that region.



(a) Grid based segmentation.

(b) Proposed segmentation algorithm.

Figure 4.11: Comparison of algorithms used to separate leg parts. Conventional grid-based segmentation is on the left and the dynamic segmentation which is based on principal eigenvectors of the upper and lower leg regions is on the right.

The axis orientation and lengths of the fitted ellipsoid are determined from the eigenvalues (d_1, d_2, d_3) and the eigenvectors (v_1, v_2, v_3) of the covariance matrix,

$$\Sigma \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}. \tag{4.2}$$

The eigenvalues correspond to the length of each of the three axes, while the eigenvector is its directional vector. The ellipsoid fitting, based on these parameters for the various position of human walking, is shown in Figure 4.12.

The orientation of the major axis and the ratios of the axis lengths are used as the features in the classification. The axis ratios are determined by,

$$r_1 = d_2/d_1, \qquad r_2 = d_3/d_1, \tag{4.3}$$

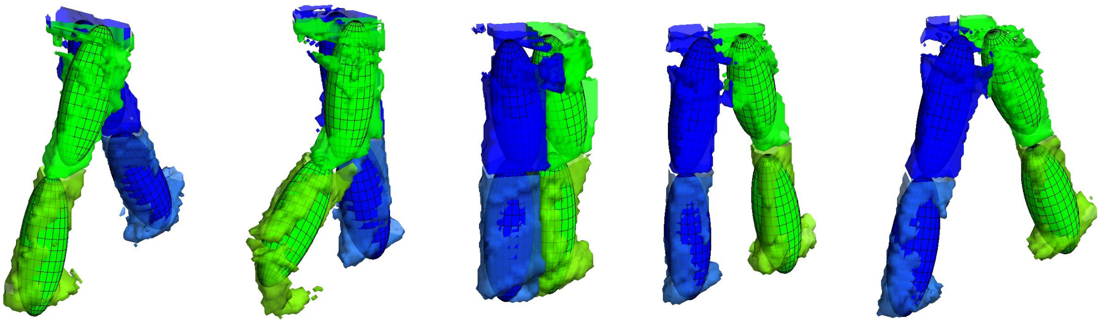


Figure 4.12: Ellipsoid fitting for different phases within a gait cycle.

while the angles used are calculated using,

$$\alpha_x = \arctan \frac{v_1 [\, 0\ 1\ 0\,]}{|v_1|}, \quad \alpha_y = \arctan \frac{v_1 [\, 0\ 0\ 1\,]}{|v_1|}, \quad \alpha_z = \arctan \frac{v_1 [\, 1\ 0\ 0\,]}{|v_1|}, \tag{4.4}$$

assuming d_1 and v_1 are the eigenvalue and eigenvector corresponding to the major axis. From this, five features are there for each region, for a total of 20 extracted features for each frame.

Gait Cycle Segmentation in 3D

Since the left and right legs are segmented separately in the proposed approach, it is easy to distinguish the left and right strides and separate each gait cycle. The beginning of a gait cycle is defined as when the legs are at greatest separation and the right leg is in front. The stride length is used to segment the gait cycles as it forms a reasonably clean, cyclic signal, with peaks corresponding to the extreme strike points of each leg. This signal is smoothed using a median filter and peaks are identified using the nearest neighbour extreme minimum computation algorithm used by Sarkar*et al.* [67].

The y-axis angles of the upper legs are used to determine the location of the left and right legs. A logical mask is created by comparing these angles, with a smoothing window applied to remove any noise. The final mask is then used to separate the peaks as belonging to the left or right legs (Figure 4.13).

4.3.3 Experiments and Results

The extracted ellipsoid parameters vary throughout the gait cycle and form a signal. To obtain the final feature vector, harmonic components from these signals are extracted.

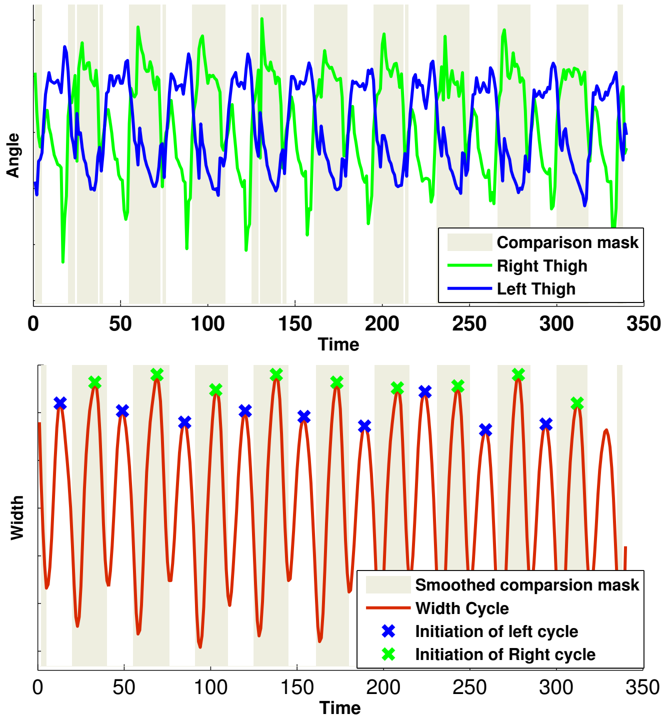


Figure 4.13: Gait cycle segmentation in 3D.

For each ellipsoid parameter, a discrete fourier transform is applied over a single gait cycle. By examining different combination of these fourier components, the first three have shown higher inter-subject variability than the intra-subject. Thus, these first three components are used as the final features in the form of complex pairs in the classification. With the first three harmonic feature for the twenty parameters, the length of the resultant feature vector for each gait cycle becomes sixty.

Before extracting the Fourier harmonics, the features are scaled such that they range between 0 and 1 in the gallery set. This scaling is then similarly applied to incoming probe sequences.

The distance, d_{ij} between i^{th} probe cycle j^{th} gallery cycle is determined by computing the Euclidean distance between the gait cycles' feature vectors (F),

$$d_{ij} = \sqrt{\sum \left|F_i - F_j\right|^2}, \tag{4.5}$$

remembering that the feature values are complex.

To determine the distance between two sequences, each of which is composed of multiple gait cycles, the algorithm proposed by Boulgouris *et al.* [124] is used as explained in Section 3.3.4 in Chapter 3.

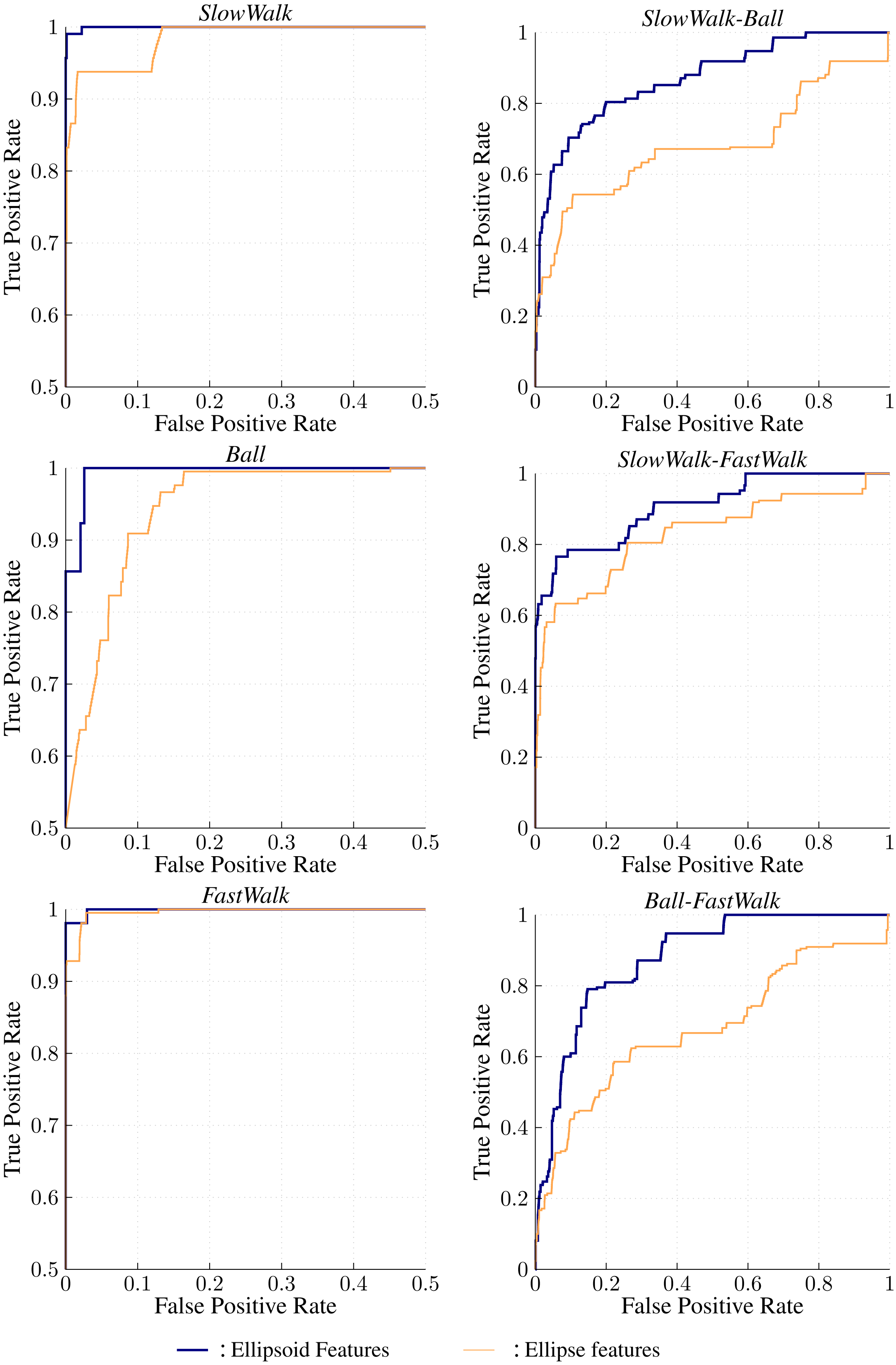


Figure 4.14: Comparison of ellipsoidal and elliptical features. ROC curves in left and right columns show intra and inter class experiments on the CMU MoBo database.

Experiment	3D-Ellipsoid Fitting	2D-Ellipse Fitting
sw	100	93.8
bl	100	91.0
fw	100	99.5
sw-bl	70.5	50.5
sw-fw	78.6	63.3
bl-fw	61.0	42.4

Table 4.1: Comparison 3D ellipsoid fitting with its baseline, 2D ellipse fitting. Verification rates at FAR of 10% are reported on the CMU MoBo database.

The performance of the algorithm is evaluated using the CMU MoBo database [31]. Both intra-class classification using slow walk, fast walk, and walking while carrying a ball, and inter-class tests between slow walk-ball, slow walk- fast walk and ball-fast walk are tested by following the protocols distributed with the database.

Similarity distance values are computed between each probe and gallery subject and performances are evaluated using ROC curves and compared with the implemented baseline algorithm in 2D. Results are shown in Figure 4.14.

As expected, intra-class tests resulted in high classification rates for both approaches. However, the proposed algorithm outperformed the baseline, achieving a 100% verification rate at a false alarm rate of 10% for all cases (see Table 4.1).

For the inter-class test cases, the proposed algorithm significantly outperforms the baseline. From Figure 4.14, it can be seen that the proposed algorithm outperforms the baseline at all operating points. At an operating point of 10% FAR, between 15-20% improvement in verification rate is achieved. This improvement can be directly attributed to the algorithm's ability to bypass the problem of self-occlusion, which hampers the original 2D method's ability to accurately model the gait dynamics.

4.4 3D-Appearance-based Gait Recognition

The model-based investigation on the CMU MoBo database shows that though it improves the recognition performance by 10-20% more than the 2D counterpart, it is lower compared to the recent appearance-based techniques such as gait energy images (GEI). However, the only disadvantage is that it can work only in a particular view and there will be significant performance degradation when view angles change. Therefore, the advantages of extending the GEI technique to 3D are explored.

The 3D appearance-based gait feature extraction method has been proposed based on the 2D gait feature, the gait energy images [98]. In this 3D approach, instead of temporally averaging segmented silhouettes, reconstructed voxel volumes are used. The resulting averaged volume is termed as the Gait Energy Volume, or GEV. To benchmark the recognition based on GEV feature, recognition performance of side-view GEI, frontal GEI, concatenated multi-view GEI are computed and experiments are carried out on the CMU MoBo database [31].

4.4.1 GEV

The proposed gait energy volumes have many advantages over gait energy images, simply by virtue of working in a three dimensional space. This circumvents the issue of view dependency, as well as having no pose ambiguity (e.g. left-right limbs), no self-occlusion, and allowing easier segmentation of unwanted regions (e.g. hand movements). However, it is not without its disadvantages, most notably the more complex hardware setup, making it impractical for many applications.

Derived from GEIs, the construction of the GEV follows a similar process to its 2D counterpart. Binary voxel volumes, analogous to silhouettes, are spatially aligned and averaged over a gait cycle as follows,

$$\mathrm{GEV}(k) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{V}(t), \tag{4.6}$$

where n is the number of frames in the k^{th} gait cycle and V is the aligned voxel volumes. An example of binary volume and corresponding GEV is shown in Figure 4.15.

4.4.2 Multi-view Gait Energy Images

As explained in Section 3.3, the implemented baseline, the GEI, is used to bench mark our proposed 3D technique. In addition to that, GEIs generated from the multiple-views are compared to ensure the effective advantage of 3D reconstruction rather than using them as raw 2D images.

The proposed GEV requires six camera views of a subject in order to achieve a complete volume reconstruction. This gives it an advantage over the GEI in benchmarks, as it has access to information not available to the GEI. A simple multi-view GEI-based algorithm is implemented in order to demonstrate the advantages of extract-

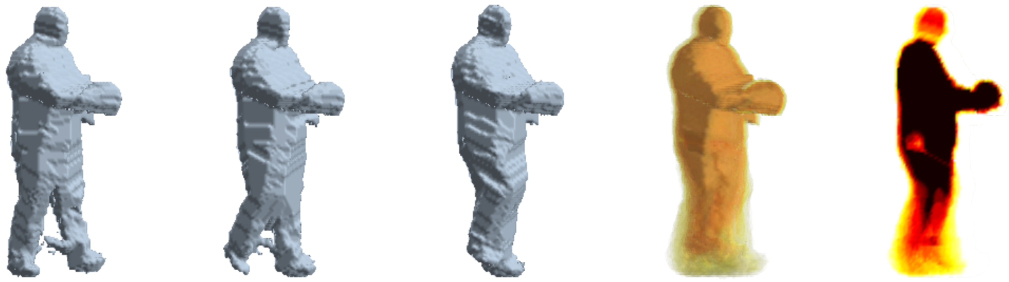


Figure 4.15: Gait energy volume. Binary voxel volumes and the GEV. A cross-sectional slice of the GEV is also shown.

ing gait energy features in 3D. In this algorithm, the GEI has been computed on the same camera views that have been used to construct the voxel volumes of the GEV. To combine the individual views, the feature vectors are extracted from each view's GEI and simply concatenated into a single super vector as shown in Figure 4.16

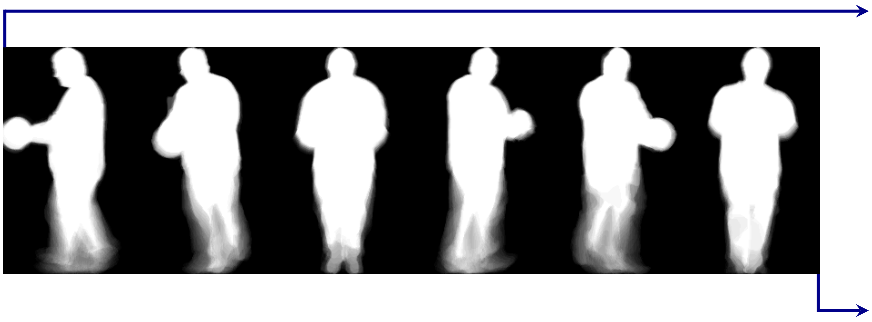


Figure 4.16: Concatenation of multi-view GEI.

The pixel values in the GEI are assembled into a feature vector, to which principal component analysis (PCA) and multiple discriminant analysis (MDA) are applied before classification as similar to the method explained in Section 3.3.3).

4.4.3 Experiments and Results

Experimental results of the implemented GEI-based and GEV-based methods are obtained using a similar method as explained for elliptical-based features. The performance of the baseline method and the proposed GEV approach are compared using ROC curves and results are shown in Figure 4.17.

Both GEV and GEI give 100% recognition rate for the intra-class experiments.

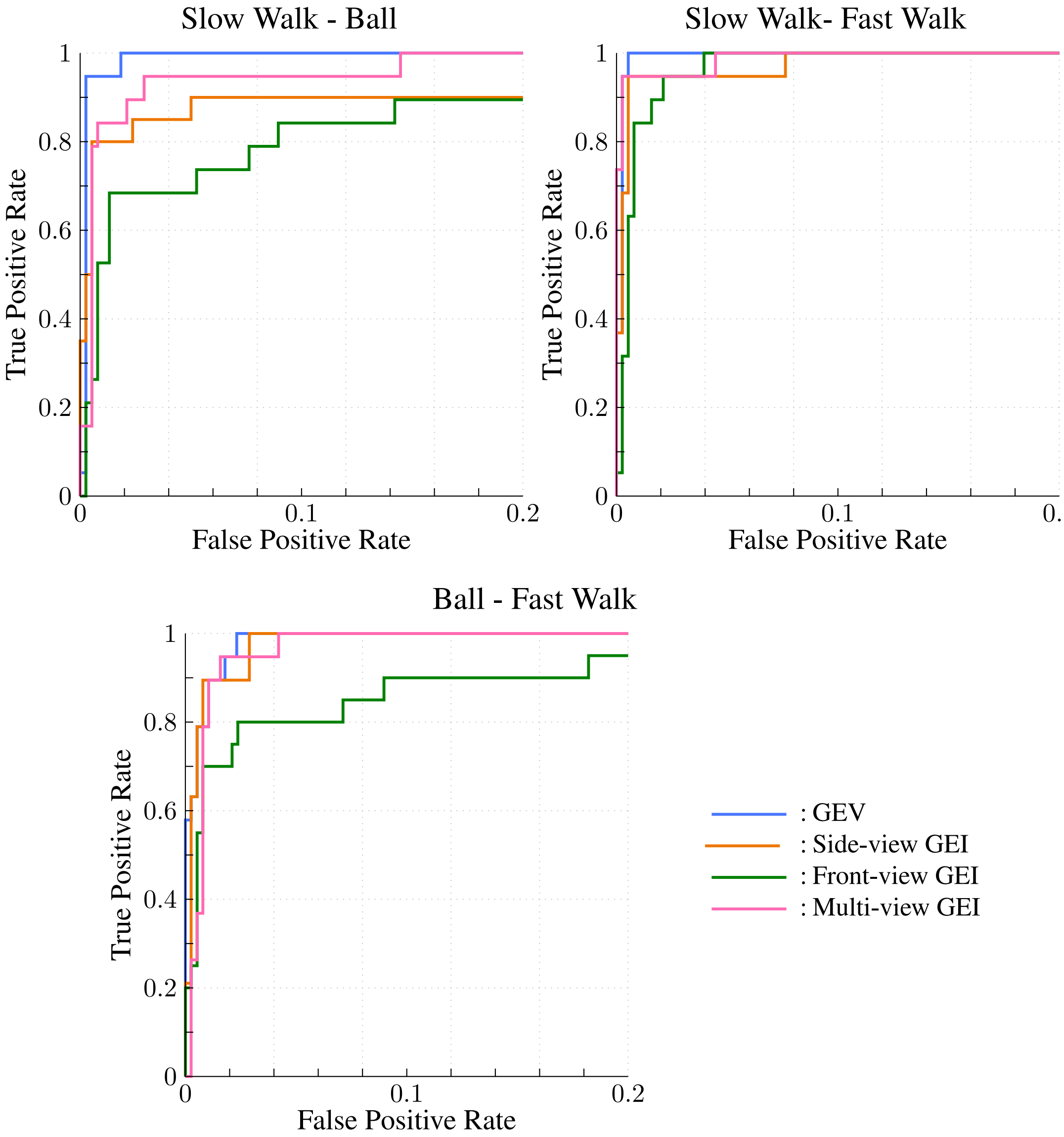


Figure 4.17: ROC curves for the evaluation of GEV and GEIs.

However, it can be seen GEV performs better and improves the results by 15% moire than other GEI techniques in inter-class experiments when the true positive rate is compared at false alarm rate of 3%. This is mostly because GEV can be segmented precisely in a 3D environment for irrelevant movements such as occlusion of moving hands in the leg region.

From the recognition performance, it can be observed that frontal GEI performs comparatively lower in all the three cases, particularly when carrying a ball. It shows that 2D frontal fails to gather all the gait dynamics due to the self- occlusion. When carrying a ball in front, silhouettes' inter-frame registration fails due to the change of centre of mass. This also significantly degrades recognition performance.

Though incorporating information from all the available views contributes to the recognition performance, it gets boosted when it is performed in 3D. This is because of better representation of static and dynamic parameters of gait in 3D. In 3D, each leg is separated and makes its individual contribution to the gait representation without occlusion. Working in 3D also helps to reduce the static noises from segmentation since it matches the voxels' projection in each corresponding view. All of these advantages confirm that the proposed method contributes significantly to solve the basic issues in underlying appearance-based techniques while, showing state-of-the art performance in CMU MoBo database.

4.5 Summary

In this chapter, incorporating multi-view data for better gait recognition has been explored by moving to 3D and two novel techniques have been proposed. With the proposed methods, improved gait recognition has been achieved by providing solutions to the initial issues in underlying techniques.

Since both model-based and appearance-based approaches have their own strengths and weaknesses, both of them have been investigated in 3D. Ellipsoidal fitting based on the eigenvalue decomposition method has been proposed as a quasi model-based technique that overcomes the issues of self-occlusion and camera-view dependency in most of the 2D model-based existing techniques. In addition to that, by using richer gait features, improvement of $10-15\%$ verification rate at false alarm of 3% is achieved over its 2D-baseline.

The proposed appearance-based method, GEV, performs state-of-the art recognition rates on CMU MoBo database by incorporating more accurate static and appearance gait parameters. The recognition performance of this method is comparatively higher than the model-based techniques, even $10 - 15\%$ higher than the proposed 3D-ellipsoidal model-based technique. This is because of missing static appearance in model-based techniques and the improper model fitting due to noisy silhouettes/volume.

Based on these observations, remaining research in this thesis will be focused on appearance-based techniques and finding solutions for the fundamental issues on them particularly for appearance-changes.

Another finding from the analysis in this chapter is that frontal gait recognition comparatively performs lower due to the lack of gait dynamics in frontal 2D because of self-occlusion. It needs to be noted that the proposed methods in this chapter require multi-view setup. In following chapters, the solutions for these issues will be addressed.

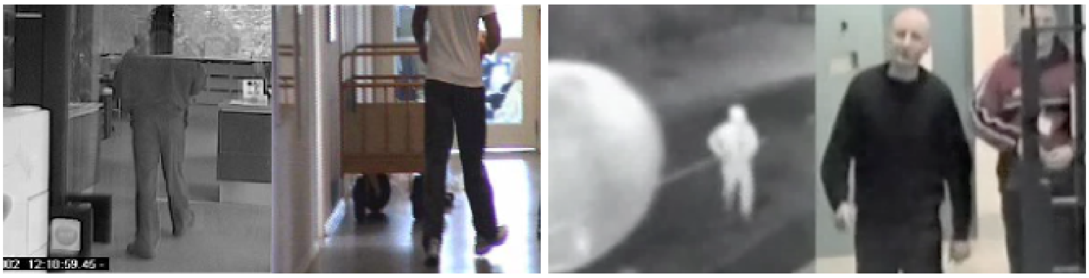
Chapter 5

Frontal Gait Recognition Using Depth Images

5.1 Introduction

Most of the gait recognition approaches in literature analyse side-view gait, for example, walking on a plane parallel to a camera. This is because most of the gait dynamics can be gathered as legs and hands extend to their maximum. However, it has been shown that in most of the forensic usages and security applications it is hard to acquire footages with side-view orientations [138]. Since CCTV cameras are usually placed on the top corners of the buildings, the subject's pose is generally captured from an upper frontal-view. There are a number of forensic analyses which have been shown in literature [138], using frontal-gait footages from CCTV cameras for criminal investigation like the case of the bank robber in Noerager [145] and the burglar in Lancashire (United Kingdom) [146]. Figure 5.1 shows example images from CCTV footage and the images captured in police stations that were used in these criminal investigations.

Another potential merit of frontal-view is that it only requires smaller physical space than the space needed in the commonly used side-view. For instance, this could be where an individual needs to verify his identity to enter a building, pass through immigration checkpoints or access any facility. In these situations, people need to line up through the narrow space where the cameras/sensors can be placed. However, to capture 8m of walking distance in side-view, camera distance of at-least 9m is required while in frontal view only the corridor type of space is adequate to capture required gait cycle as shown in figure 5.3 [147]. This potential merit explains the requirement of frontal-view in portal-based security authentication applications.



(a) (b)

Figure 5.1: CCTV footage of frontal gait recognition used in forensic analysis. Image on left from cctv and right from later in police station. (a) Burglar in Lancashire [146] and (b) the bank robber in Noerager [145]

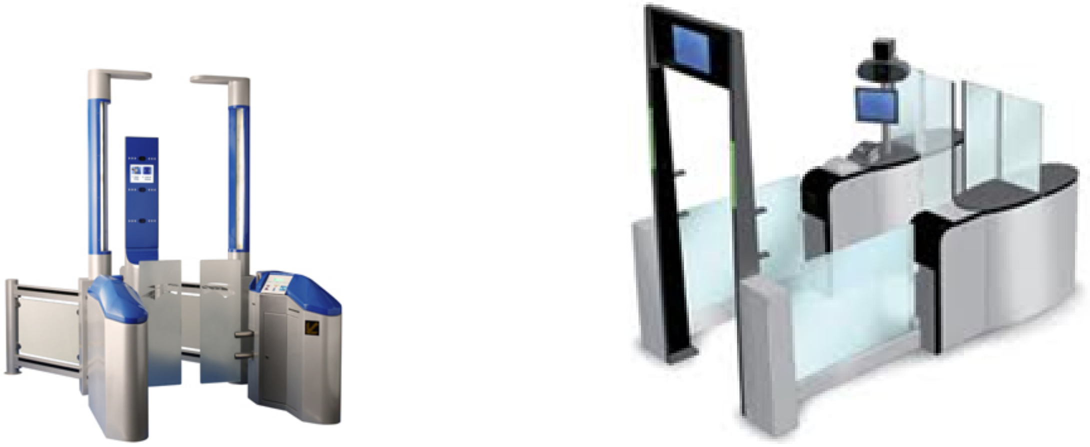


Figure 5.2: Examples of biometric-portals,equipped with face and iris based biometric authentication systems [148].

Smart-gate gives eligible travellers the option to self-process through passport control [148]. Figure 5.2 shows an example of a smart gate system used in Australian airports that uses an iris and face-based biometric authentication system. The main limitation of these iris and/or face-based authentications in smart gates, is that they need individuals captured in near-field frontal-view, where the gait-based system doesn't require this. Therefore, the gait, iris, and face-based coupled system is ideal for providing robust, near and distance field biometric authentication of an individual. However, coupling gait with these existing implementations required to place gait sensors/cameras where frontal view can be captured.

Capturing gait in fronto-parallel has its own limitation. It needs to compensate for the looming effect as well as the possibility of self-occlusion happening with legs, body and hands. Only smaller gait dynamics can be gathered on 2D image data. This is

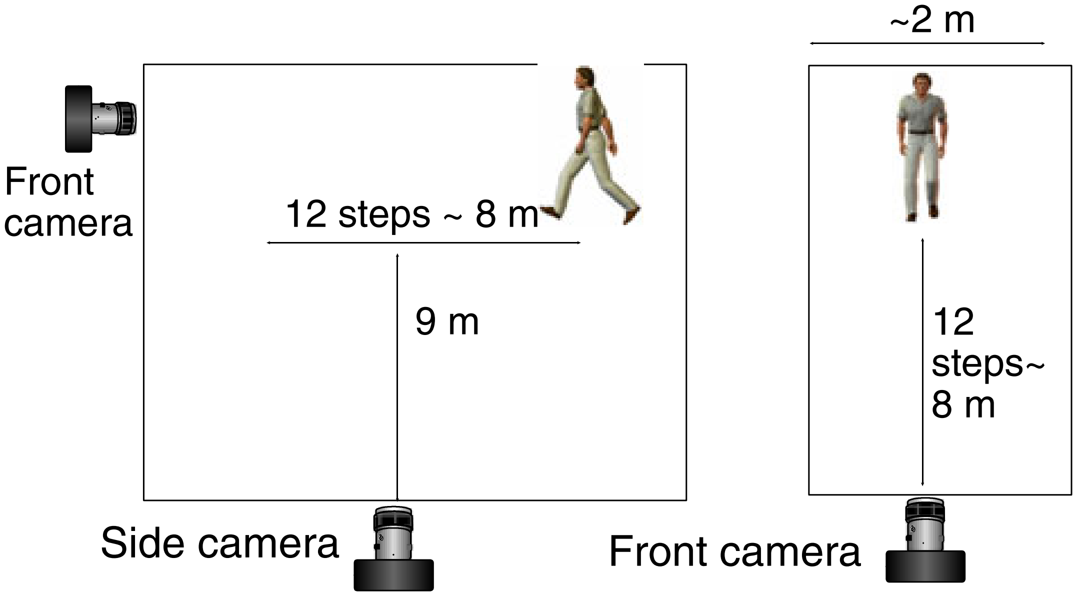


Figure 5.3: Spatial requirement to capture gait in side-view and front-view.

also evidenced from the poorer recognition performance reported on the CMU MoBo database in Section 4.4.3 of Chapter 4.

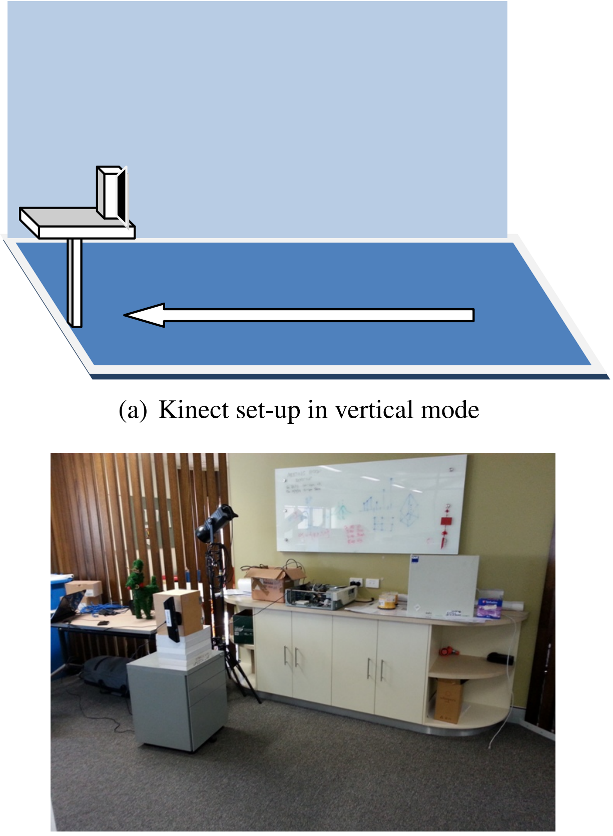
In Chapter 4 appearance-based methods significantly outperform the model-based and working in 3D boost the performance. Based on this conclusion, 3D gait appearance-based features need to computed from frontal-view for robust gait recognition. However, 3D reconstruction required multiple-camera set up that impractical in concise space as discussed earlier. An alternative to acquiring this 3D data would be to use depth sensing device. Frontal based depth has the advantage of being able to capture essentially all characteristics of gait from a single viewpoint without the issue of self occlusion. These depth images can be easily acquired using devices such as Microsoft Kinect [149].

5.2 Gait Recognition Using Frontal Depth Images

Promising results of GEV in full 3D reconstructed voxel volumes in Chapter 4 motivates to extend that to apply on frontal depth. As only the front surface of the subject is visible from depth images, only a partial volume reconstruction is possible. The voxel model is created by taking the frontal surface reconstruction and filling to the back of the defined voxel space along the 'depth' axis. Based on the assumption that only frontal leg information is adequate enough to distinguish an individual, frontal GEV is generated as frontal gait features using these backfilled frontal volumes by following the same procedure as for full 3D GEV.

5.2.1 Frontal Depth Gait Database

To facilitate the research on frontal gait using depth images, and test the proposed 3D GEV algorithm on the real-world application scenario, Frontal Depth Gait Dataset, DGD, is proposed. The dataset is captured at approximately 30 fps using Microsoft Kinect [149]. Colour video was also recorded but was not used. To facilitate greater field of view, Kinect was placed in vertical mode as shown in Figure 5.4. Each walking sequence for an individual covers an average of 2 - 3 gait cycles, though only about two cycles are useful due to limitations in depth resolution.



(a) Kinect set-up in vertical mode

(b) Walking Corridor

Figure 5.4: Kinect set-up for data capture.

The frontal DGD database consists of 35 subjects walking towards the camera under six different walking conditions: normal walk (*nw*), fast walk (*fw*), back carrying (*bc*), side carrying (*sc*), front carrying (*fc*) and no shoes (*ns*). Multiple sequences are captured for each subject and walking condition, with five each for *nw* and *fw*, four for *ns*, and two for the others.

Example images from each condition are shown in Figure 5.5. RGB and depth images are freely available for research purposes at `http://df.arcs.org.au/quickshare/87362d2fcd6f1503/SAIVT-DGD.tar.gz`.

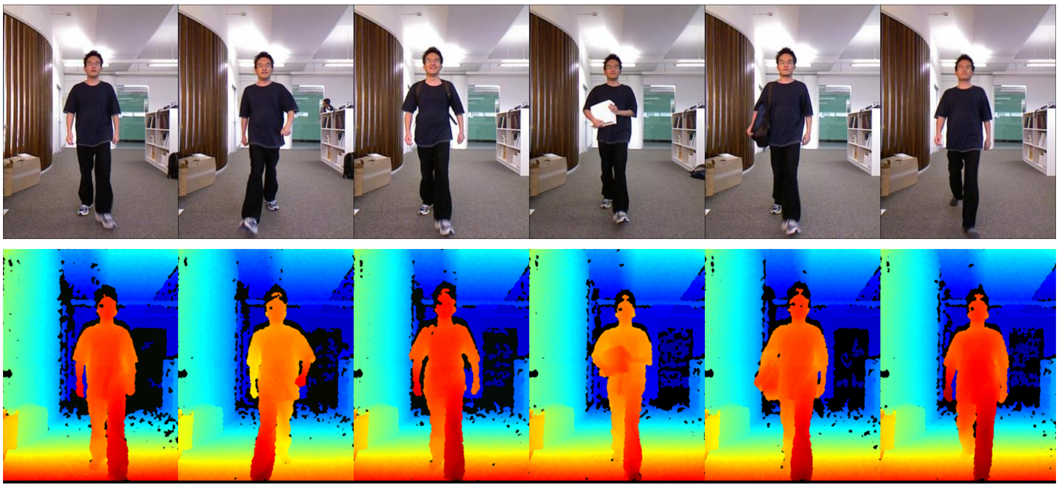


Figure 5.5: RGB and depth images from depth gait database left to right shows each condition *nw, fw, bc, sc, fc, ns*).

5.2.2 Frontal Volume Reconstruction

Initial step to create frontal volume from depth data is transferring depth value, $d_{r,c}$, to world coordinate, (w_z) in meters using Microsoft Kinect's [149] intrinsic parameters as follows,

$$w_z = 0.075000003 * 587 * 8./(1090 - d_{r,c}). \tag{5.1}$$

Kinect depth camera is then calibrated to compute its intrinsic parameters, focal length (f_r, f_c) and camera center (c_r, c_c). Using these intrinsic parameters of the depth image row, column indexes of the depth image are converted to world coordinates as follow,

$$\begin{aligned} w_y &= (r_g - c_r). * w_d/f_r, \\ w_x &= (c_g - c_c). * w_d/f_c, \end{aligned} \tag{5.2}$$

where r_g and c_g are mesh-grid values computed using row and column indices of the depth image, r and c and (w_x, w_y, w_z) that represent the frontal surface in world coordinates as shown in Figure 5.6. Transformation to world coordinates makes segmenta-

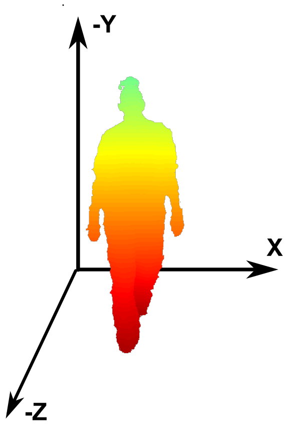


Figure 5.6: Axis convention for kinect depth data.

tion of silhouettes easier, since that can be done by extracting depth pixels between the floor, ceiling and wall boundaries. Left, right, front, back and top segmentation boundaries are defined by independent inequality, vertical or horizontal planes, considering the worse-case scenarios as follow,

$$\text{Silhouette surface } = W(x, y, z), \tag{5.3}$$

$$\text{where,}$$

$$z_{min} \geq z \leq z_{max},$$

$$x_{min} \geq x \leq x_{max},$$

$$y \leq y_{min}.$$

Since there are horizontal variations in the floor plane, due to the changes in Kinect orientation, floor plane parallel to XZ is not accurate enough to separate the floor pixels from the human silhouette. Therefore, the accurate floor plane equation has been computed by selecting three points on the ground as shown in Figure 5.7. These selected points are used to solve the plane equation $ax + by + cz + d = 0$.

Figure 5.8 illustrates the plane segmentation in each x, y, z direction using corresponding colour map for better visualisation and Figure 5.9 explains the overall segmentation process to extract the human silhouette surface in 3D.

To construct the voxel-volume of the extracted surface, new axis convention is defined as ($z = -z$ & $y = -y$) and world coordinates of the surfaces are projected to

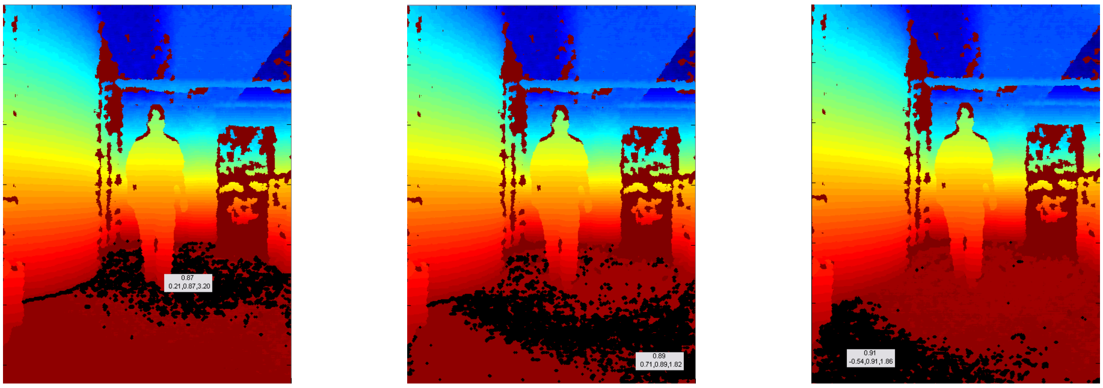
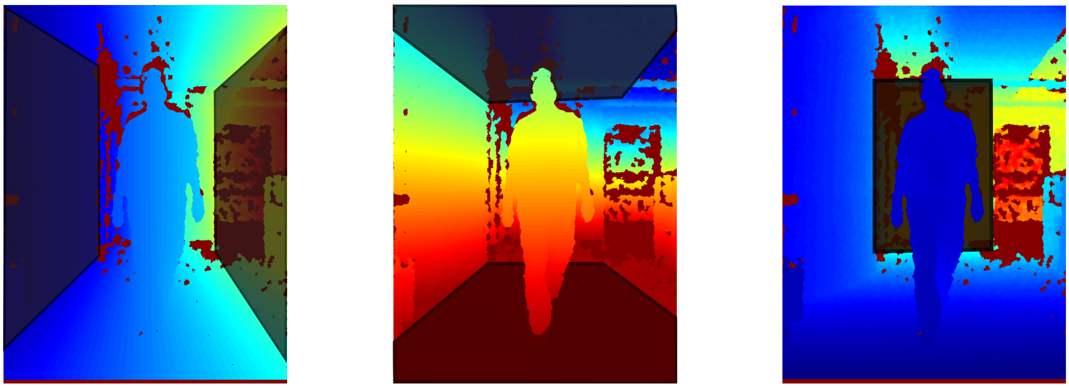


Figure 5.7: Point selection for ground plane computation.



Segmentation in x direction

Segmentation in y direction

Segmentation in z direction

Figure 5.8: Segmentation planes on 3D world coordinates. Colour maps are used for better visualisation in each direction.

voxel coordinates, (x_v, y_v, z_v), using the voxel resolution v_r as below,

$$\begin{aligned} x_v &= (x - \min(x)) * v_r + 1, \\ y_v &= (y - \min(y)) * v_r + 1, \\ z_v &= (z - \min(z)) * v_r + 1. \end{aligned} \tag{5.4}$$

Mean, (c_{x_v}, c_{y_v}, c_{z_v}) is used to align the voxels. Alignment based on mean value makes temporal movements of the subject almost based on their hips and preserve the height information also there is a chance to reduce the peak noises. To do this voxel volume, vol_{d_x,d_y,d_z} is defined with the size of d_x, d_y, d_z and computed voxels are aligned with

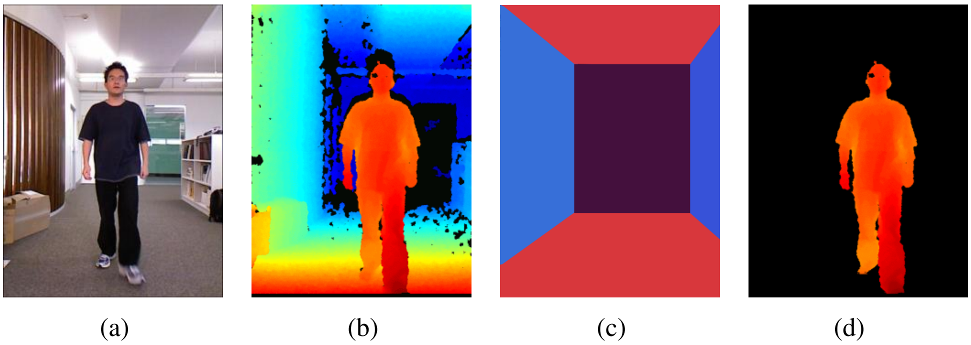


Figure 5.9: Overview of plane segmentation process. (a) Frontal RGB image (b) depth image (c) segmentation planes and (d) segmented depth image.

the centre of the volume as follow,

$$\begin{aligned} x_v &= \lceil x_v - (c_{x_v} - \frac{d_x}{2}) \rceil, \\ y_v &= \lceil y_v - (c_{y_v} - \frac{d_y}{2}) \rceil, \\ z_v &= \lceil z_v - (c_{z_v} - \frac{d_z}{2}) \rceil. \end{aligned} \tag{5.5}$$

Voxels beyond the size of the pre-defined voxel volumes are discarded and holes on the voxels are interpolated by approximation using sparse linear algebra and discretisations of partial differential equations. Frontal voxel volume is then computed by back-filling the surface to the backside of the pre-defined volume as follows,

$$vol(x_v, y_v, 1 : z_v) = 1. \tag{5.6}$$

This backfilling of surface is better illustrated in Figure 5.10.

5.2.3 Frontal Gait Energy Volume

Frontal gait energy volume is computed by averaging the frontal volumes over their gait cycle as similar to the method used to compute the GEV (See Section 4.4.1). Stride length in depth direction (Z axis direction in Figure 5.10) is used to segment gait cycles and extreme striking points of each legs are detected by computing the peaks of this stride length signal. Figure 5.11 shows examples of constructed volume, as well as a computed frontal GEV with its slice view.

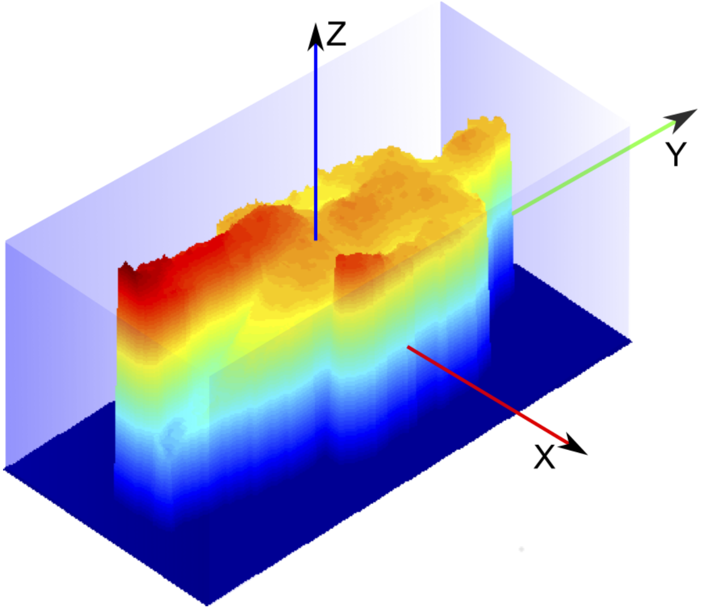


Figure 5.10: Computed backfilled frontal volume.

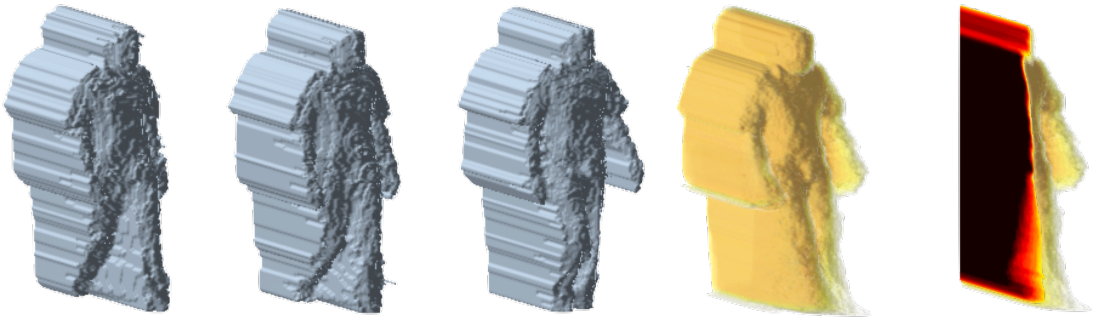


Figure 5.11: Frontal gait energy volume. Frontal backfilled voxel volumes and the frontal GEV. A cross-sectional slice of the frontal GEV is also shown.

5.3 Gait Recognition on Synthesised Depth Volumes

Since there are no frontal-depth gait databases evaluated in literature, a depth reconstruction is simulated using multi-view camera data, CMU MoBo dataset. This allows us to benchmark our results against existing gait algorithms on a common database, as well as allowing us to quantify what effects, if any, are produced by removing the back-face of the subject model.

To generate the frontal voxel-volume, the front surface of the full voxel model is found (as described in Section 4.2) and filled to the back of the volume boundary as below,

$$\mathrm{vol}_{(x,y,1:z)} = 1. \tag{5.7}$$

where (x, y, z) corresponds to the coordinates of the voxels on the silhouette surface. Generated frontal volumes are again aligned to the centre and frontal GEV is computed as described in Section 4.4.1. Figure 5.12 compares the examples of synthesised depth volumes and corresponding frontal GEV with the volumes and GEV generated using full 3D.

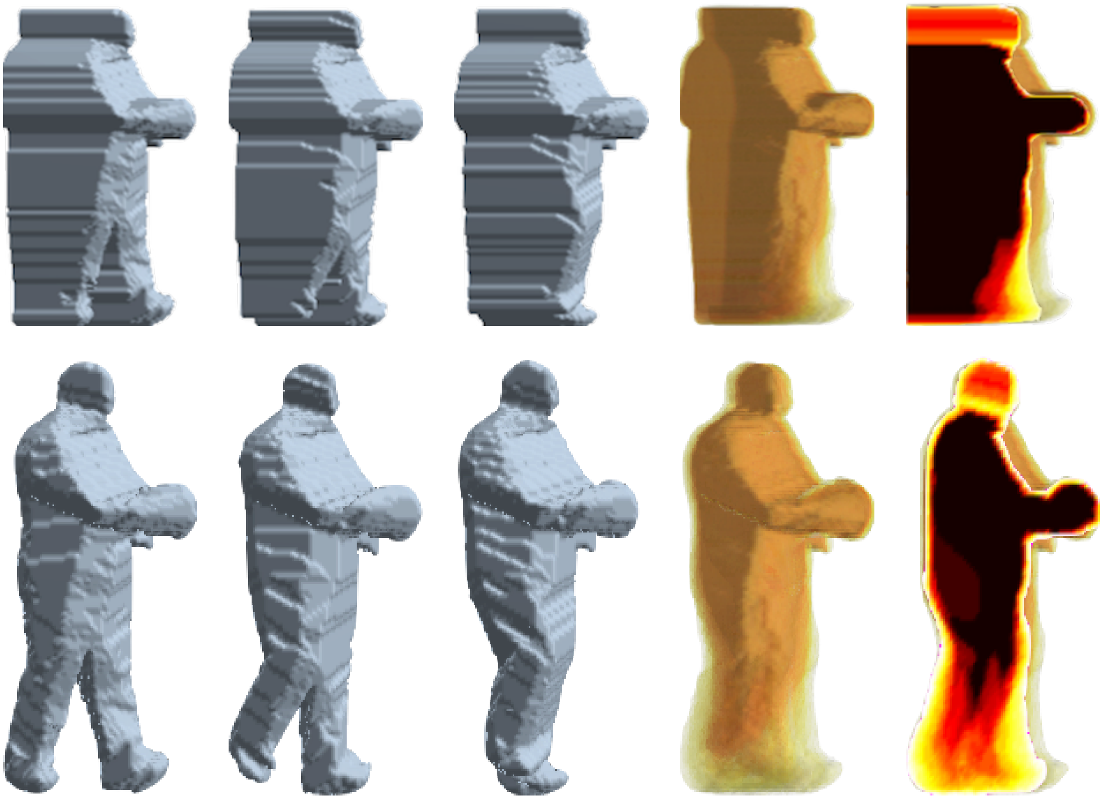


Figure 5.12: Frontal gait energy volume on synthesised frontal volumes and full 3D volumes (Frontal volumes and, the frontal GEV and its cross-sectional slice, using synthesised data (top) and full 3D data (bottom)).

5.4 Experiments and Results

There are two sets of experiments that are carried out to evaluate the proposed frontal GEV feature. In the first of these experiments, recognition performance of frontal GEV from depth images is evaluated.

Second set of experiments are carried out to evaluate the effect of removing back leg information on frontal GEV by comparing the recognition results of synthesised frontal GEV with the full 3D GEV.

For both experiments, computed frontal GEV features are transformed to a discriminative energetic component domain by using the PCA and MDA basis that is learned from the training data. Distance between gallery and probe gait cycles is computed using Euclidean distance and single similarity distance score for each probe and gallery subject is computed, as illustrated in Section 3.3.4 in Chapter 3.

5.4.1 Depth Frontal Volumes

The experiments on depth frontal volumes are evaluated on the in-house DGD database by allocating three out of the five *nw* as gallery cycles and the rest as probe cycles for intra-class tests (*nw-nw*). In inter-class tests, all extracted gait cycles in *fw*, *sc*, *fc*, *bc* and *ns* are allocated as the probe and tested with the available five gallery gait cycles,*nw*.

Figure 5.13 shows ROC curves for the experiments carried out on the DGD database. It has been noted that recognition performance of more than 99% was achieved at FAR 1% for the experiments, other than frontal carrying with normal walk. Lower performance in *fc-nw* may be due to the changed appearance of frontal surface, due to the carried goods.

5.4.2 Synthesised Frontal Volume

The experiments evaluate the frontal GEVs extracted through synthesised ‘depth’ construction with the GEVs generated from the visual hull of the multi-view silhouettes from CMU MoBo database. Recognition performances are compared with the baselines, the GEIs from front-view, side-view and multi-view, explained in Section 4.4 of Chapter 4.

Both intra and inter-class cases are considered and Receiver Operating Curves (ROC) are used to compare the results. In the intra-class cases, the video sequence is split in half, with the gait cycles in the first half used as the gallery and the second half used as the probe. Intra-class recognition performance of all algorithms is 100%, without any false alarms. For inter-class tests, the full sequence is used as either the gallery or probe, in the combination shown in Table 3.2 in Chapter 3.

ROC curves for the inter-class experiments are shown in Figure 5.14. From the results, it can be seen that our GEV approaches outperform the GEI. This includes the multi-view GEI, showing that it is worth-while performing the simple volume reconstruction and working in a 3D space, given multi-view data. The GEV applied to the

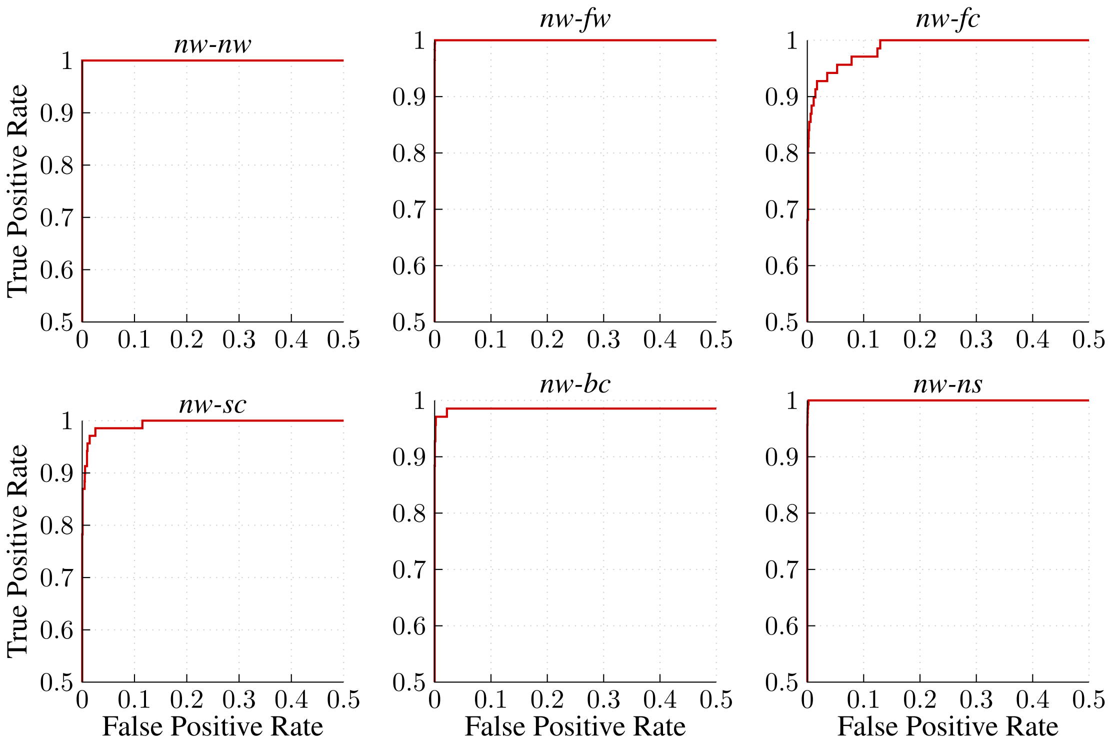


Figure 5.13: ROC curves for the tests on the DGD database using frontal depth GEV features

synthesised depth reconstruction performs similar or greater to the full volume GEV (average recognition rate of 100% at FAR at 1% is achieved using frontal GEV where, only 94% is achieved using full volume GEV). This is likely due to essentially all gait characteristics being acquired from a frontal perspective, while the back filling reduces the impact of noise.

5.5 Summary

This chapter explores the solutions for effective frontal-based gait recognition to integrate distance-field biometric gait, with the existing near-field biometrics' face and iris in smart gate-based applications. A single camera solution using frontal-depth is proposed to solve the self-occlusion in frontal 2D gait features. Appearance-based 3D feature, GEV, is selected as a gait feature and computed from backfilled frontal volumes using depth images. To evaluate the algorithm, frontal depth gait database (DGD) is also developed and made available for future research to enable the new direction to frontal gait recognition. The verification results on the DGD database using the proposed algorithms shows an average of 98% at FAR of 1% with 35 subjects and

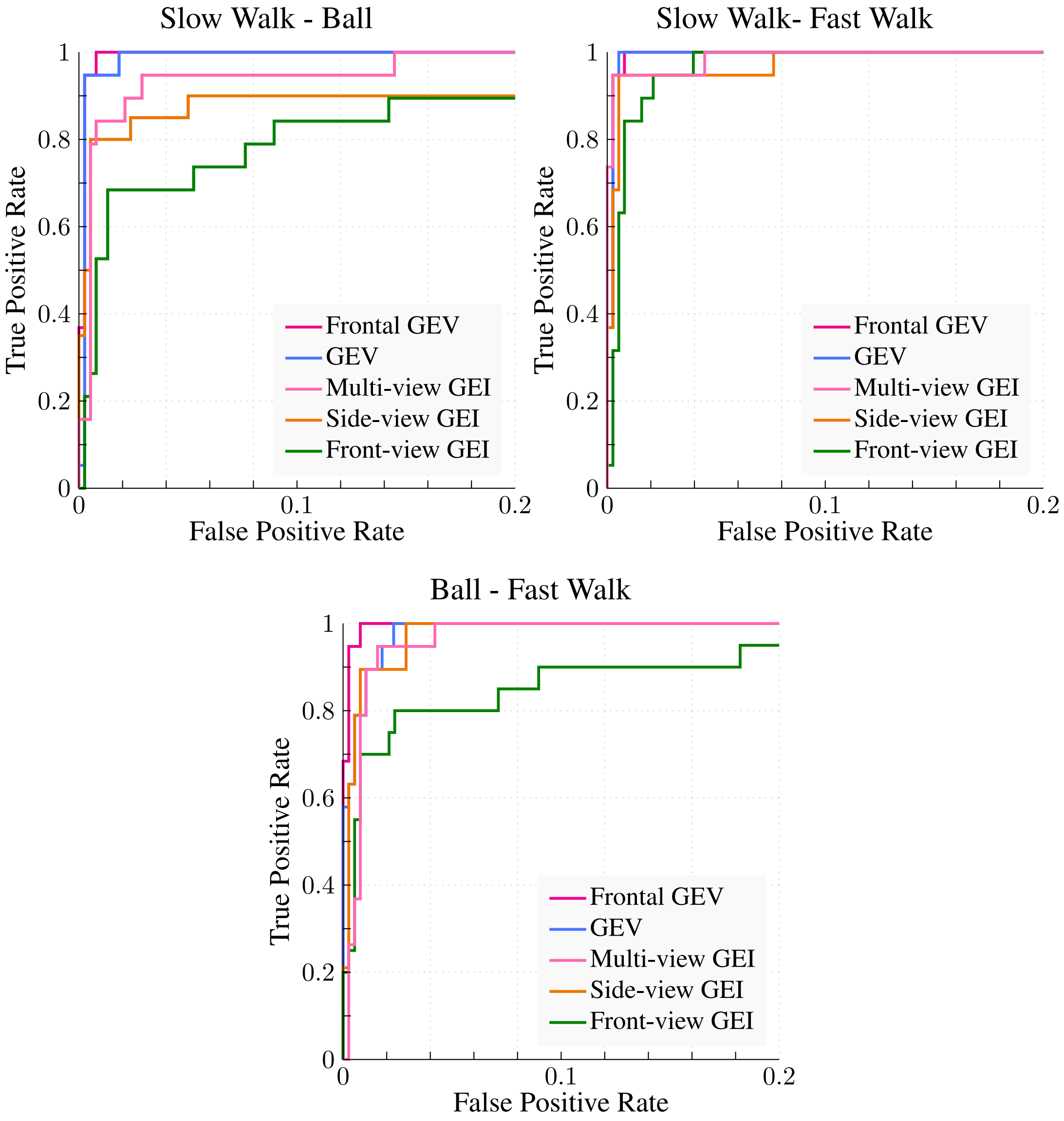


Figure 5.14: Recognition performance of the GEV on the Synthesised frontal depth data from the CMU MoBo database.

five different challenging conditions.

In addition to that, the assumption that “frontal gait has all the essential information to recognise an individual” is verified on the CMU MoBo database with improved recognition results compared to the full 3D GEV. These higher recognition rates show the potentiality of the proposed solution for the real world walk-through biometric portals.

Chapter 6

Feature Optimisation and Robust Classification

6.1 Introduction

In previous chapters, improved recognition performance has been achieved for appearance-based methods by combining robust pre-processing tasks with 3D/depth. However, these methods still struggle to perform in inter-class conditions with appearance changes. There are two main reasons for this performance degradation.

- Poor performing classifier
  Effectiveness of the classifier performs a major role in improving the recognition performance. However, in most of the popular gait recognition algorithms ([98, 101, 150]), the nearest neighbour (NN) [143] is used for classification because of its simplicity and the non-requirement of a learning model. In NN, the test sample is represented in terms of a single training sample; this is not feasible in occluded/ noisy scenarios particularly in inter-class test cases.

- Improper registration of inter-class gait features
  Registration of features in the spatial and time domain is important in matching them in the classification process. However this can be affected due to the segmentation noise and appearance changes in interclass test cases.

This chapter explores the methods to solve these two issues. At first, to address the lack of representation of test subject in NN classifier, a sparse representation-based classifier (SRC) is explored in gait context. Improvements have been shown in face recognition [52] by adopting a sparse representation-based classier, where the test subject is represented by considering all the possible contributions from within class and between class training data. In this chapter, a SRC-based classifier is explored for robust classification in the gait recognition context, together with supervised learned discriminated input feature space.

To provide the solution for improper registration in inter-class test cases, local grid-based histogram descriptors are adopted in the input feature space, since spatial sampling concepts in these algorithms can handle intra-patch misalignment between frames. The effectiveness of the existing patch-based descriptors such as a histogram of oriented gradients (HOG) [151], local binary pattern (LBP) [60] and local directional patterns (LDP) [61] are explored in the gait recognition context by applying them to GEI. We also propose a novel feature, a histogram of weighted local directions (HWLD) that combines the strengths of HOG and LDP while optimising the histogram descriptor in a gait recognition context.

6.2 SRC-based Classifier for Gait Recognition

As with many other appearance-based recognition tasks in computer vision (*e.g.* face recognition), some form of Euclidean distance-based nearest neighbour algorithm is commonly used for classification, typically after applying principal component analysis (PCA) [152], or some other dimensionality reduction technique.

Recently, sparse representation based classification [51] has become popular and has been used as an effective classification method for face [52], action recognition [54] and facial expression recognition [118].

Since initial investigation on sparse representation results shows the improvement over traditional classification methods, the applicability of SRC-based classification is investigated similar to that used in face recognition [52], and propose using a better discriminated input feature space to improve performance.

Classification in NN is performed by representing the test sample in terms of a single training sample. Another version, called nearest subspace (NS) [153], models the test subject as a linear representation of all the training samples in each class. A combination of these two techniques has been introduced, as nearest feature line (NFL) [154] classifies based on the best affine representation, in terms of a pair of training samples.

The sparse representation-based classifier can be considered a generalisation of nearest neighbour and nearest subspace [153]; it adaptively chooses the minimum number of training samples needed to represent each test sample.

The remainder of this section explains the theory behind sparse representation and its extension to perform gait-based person recognition. It also explains the improved SRC by performing global discrimination on the locally discriminated input feature space.

6.2.1 Compressive Sensing for Sparse Representation

Compressive sensing and sparse coding [51,53] are emerging fields in statistical signal processing which are widely used to compute sparse linear representations of an object for compression or reconstruction potentially using lower sampling rates than the Shannon-Nyquist bound. Compressive sensing provides the solution to form and accurately recover the compressed signal without any requirements on the number of initial samples, on the need to store the coefficients and locations. It directly acquires the signal representation without going through the intermediate stage of acquiring input samples.

The objective is to transform the observed signal $x \in \mathbb{R}^N$ to a compressed signal $y \in \mathbb{R}^M$, where $M < N$, using a measurement matrix $\phi \in \mathbb{R}^{M \times N}$ as follows,

$$y = \phi x. \tag{6.1}$$

However accurate reconstruction of signal x from y is ill-conditioned since $M < N$. To find a solution to this problem, the sparsity of the signal x is exploited by transforming x into ψ domain as $s = \psi^T x$, where ψ is a sensing matrix and s sparse coefficients of x in ψ space. If x is K sparse for some scalar $K < M \in \mathbb{R}$ with the known location of K non-zero elements then the problem can be solved with the restricted isometric property (RIP) [155] of $\Theta = \phi\psi \in \mathbb{R}^{M \times N}$ for some scalar σ as follows,

$$1 - \sigma \leq \frac{\|\Theta s\|_2}{\|s\|_2} \leq 1 + \sigma. \tag{6.2}$$

This condition ensures the sparsity of s is preserved throughout the compression. In [51, 156], it has been shown that for $\psi = I$, $M \times N$ iid Gaussian matrix, $\theta = \phi I$, have RIP with high probability, if $M \geq cK \log(NK)$ for small c.

The reconstruction part of the above compressive sensing discussion provides hints for the solution to the classification problems. Representing test object as linear measurements of the training objects will be sparse since most of the coefficients that do not correspond to the test object are zero. The remainder of this section explains the formulation of this classification problem and the solution by sparse representation.

6.2.2 Classification Using Sparse Reconstruction

A basic classification problem in human identification is correctly determining an individual to whom the test features belongs to, from the dictionary of enrolled subjects. To explore this classification problem using gait features, the similar principles that used on the face features have been followed by formulating a low dimensional subspace [157].

Let us arrange the gait features of k_i training samples from the i^{th} subject in a column matrix, $D_i = [w_{i,1}, w_{i,2}, ..., w_{i,k}] \in \mathbb{R}^{m \times k}$ and formulate the dictionary, $D = [D_1, D_2, ..., D_n]$, where $w \in \mathbb{R}^m$ is input gait feature and n is number of training subjects. If k_i is sufficiently large, then the test subject $\gamma \in \mathbb{R}^m$ can be approximately represented by a linear equation for some scalers $\alpha_{i,j} \in \mathbb{R}, j = 1, 2, ..k_i$ as below,

$$\begin{aligned} \gamma =& \alpha_{1,1} w_{i,1} + ... + \alpha i - 1, k_{i-1} w i - 1, k_{i-1} + \alpha_{i,1} w_{i,1} \\ &+ ... + \alpha_{i,1} w_{i,k_i} + \alpha_{i+1,1} w_{i+1,1} ..., \alpha_{n,k_n} w_{n,k_n}. \end{aligned} \tag{6.3}$$

For an ideal situation, the test sample γ corresponding to i^{th} subject can only be represented by the training samples of i^{th} subject and Equation 6.3 can be re-written as $\gamma = D\alpha_0$, where $\alpha_0 = [0, 0, ..., \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,k}, 0, ..., 0]^T$. However for the unknown test subject γ, the identity is claimed/verified by solving the linear equation for α as below,

$$\gamma = D\alpha. \tag{6.4}$$

Equation 6.4 is over-determined when feature dimension, m, is greater than the number of subjects, n, in dictionary, D. Then, for the full rank matrix, D unique solution , α_0, can be found by Gaussian elimination or similar methods.

For closed form classification problems (where the subject is already registered in the gallery), the number of subjects in the training dictionary is large and typically $\gamma = D\alpha$ is under determined (m¡n) even D has full rank. Though knowledge of $\gamma = D\alpha$ restricts α to an affine subspace of $\mathbb{R}^n$, the unique solution cannot be computed completely. If α is small, a least square solution can be used for α as follows,

$$\hat{\alpha_2} = \underset{\alpha:\gamma=D\alpha}{\operatorname{argmin}} \|\alpha\|_{\ell_2} = A^*(AA^*)^{-1}\gamma. \tag{6.5}$$

However, the above Least-square solution by pseudo-inverse is not satisfactory, as it is an energy-based optimisation solution and the resulting $\hat{\alpha_2}$ is dense with many non-zero coefficients. Therefore this is not the sparsest solution to identify the test subject. To solve the problem, the intuitive nature of the subspace formed by the optimisation problem is considered. If many training samples corresponding to the test subject are included in the dictionary, quite often it's possible to represent the test sample as a linear combination of that subject's training samples only. In addition to that, if the number of subjects, $n >> k$, then the solution to α is sparse. The sparsest solution of $\gamma = D\alpha$ will be the correct solution for γ.

As discussed in Section 6.2.1 and from the proposition proved in [158], if any $2S$ columns in $m \times n$ matrix, D, are linearly independent, then any S sparse signal, $\alpha \in \mathbb{R}^n$, can be reconstructed uniquely from $D\alpha$. The proof of this proposition explains that the sparsest solution of α can be computed by ℓ_0 minimisation as follows,

$$\hat{\alpha_0} = \underset{\alpha:\gamma=D\alpha}{\operatorname{argmin}} \|\alpha\|_{\ell_0}, \tag{6.6}$$

where,

$$\|\alpha\|_{\ell_0} := \sum_{i=1}^{i=n} |\alpha_i|^0 = \#\{1 \leq i \leq n : \alpha_i \neq 0\}. \tag{6.7}$$

In contrast to ℓ_2 norm minimisation, ℓ_0 minimisation is computationally difficult even for approximation and it is NP-hard in general because it is not a convex optimisation problem. However, compressive sensing states that the sparsest solution can be found for the under-determined system by using ℓ_1 norm minimisation as follows,

$$\hat{\alpha_1} = \underset{\alpha:\gamma=D\alpha}{\operatorname{argmin}} \|\alpha\|_{\ell_1}. \tag{6.8}$$

The above equation is a convex optimisation problem and it can be solved by linear programming techniques such as [159].

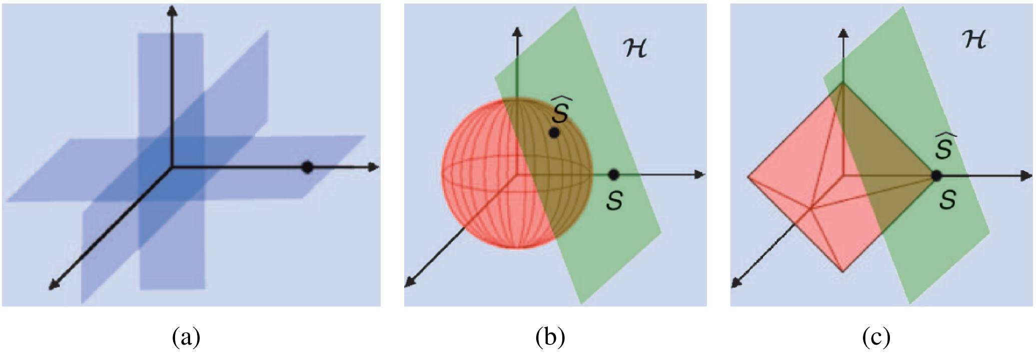


Figure 6.1: 3D geometric explanation of ℓ_2 and ℓ_1 norms ([160]. (a): Two sparse vectors in $\mathbb{R}^3$, (b): Solution $\hat{S}$ where ℓ_2 ball hits the translated null space (green) and (c): Solution $\hat{S}$ where ℓ_1 ball hits the translated space.

6.2.3 A Geometric Explanation of the L1-norm Solution

The reason for selecting $l1$ norm rather than more common $L2$ norm is explained by how the $\ell_1$norm and ℓ_2 norm find the solution to the Equation 6.12. The geometric explanation in a three dimensional case is illustrated in Figure 6.1. The K-sparse vectors in $\mathbb{R}^N$, *i.e.* those aligned with the coordinate axes, are shown in Figure 6.1(a). These sparse vectors are translated to null-space $H = N(\theta) + s$. In null space, H, the actual solution to the test subject, S lies on the coordinate axes because of its sparsity, as shown in Figure 6.1(b).

In $\ell_2 norm$ optimisation, the ℓ_2 ball is grown until it touches the hyper plane for the null space, H, and finds the nearest point from the origin as a solution, $\hat{S}$. By referring to Figure 6.1(b), $\hat{S}$ is neither sparse, or close to the correct solution, S. However, points of the ℓ_1 ball are aligned with the coordinate axes and when it is grown, it touches the hyper plane of translated null space at a point near to the coordinate axes as shown in Figure 6.1(c). The solution is sparse and at the same time near to the actual point, S.

6.2.4 MDA with SRC

The objective of SRC is to represent the test subject with a sparse combination of the learned dictionary subjects. However, it is impossible to achieve the exact representation of a test subject by a sparse superposition with only non-zero coefficients of the relevant subject because of appearance changes and other external varying factors (*e.g.* the carrying of goods). The assumption that SRC can work effectively regardless of the feature space [51] needs to be revised in this scenario. The ℓ_1 norm optimisation as explained in Equation 6.8 tries to find the solution globally, and hence fails to identify the similarities and differentiating attributes within and between subjects by local analysis, even though the local subject labels are available.

To solve this issue, multiple discriminant analysis (MDA) is applied to analyse the class-labelled data for intra-class (within the subjects) similarities and inter-class (between the subjects) dissimilarities. To extract the most discriminant features from the projected feature vectors, MDA is applied to learn the transformation matrix, T_{mda}. Transforming the feature vector on the leaned basis domain maximises the ratio of the between-subject scatter matrix to the within-subject scatter matrix. This is done by computing the generalised eigenvector that corresponds to the largest eigen-values of the within-subject and between-subject scatter matrices. The dimension of the projected feature space has been chosen as one less than the number of subjects, as explained in [98].

Each column vector in the dictionary and the feature vector of the test subject has been transformed using T_{MDA} as follows,

$$\hat{w} = T'_{mda} \times w, \tag{6.9}$$

where $\hat{w}$ is the final feature vector, after MDA transforms have been applied. The above locally discriminated features form the new dictionary, $\hat{D}$, and the transformed feature vector of a test subject, $\hat{\gamma}$, and ℓ_1 minimisation problem becomes as below,

$$\hat{\alpha_1} = \underset{\hat{\alpha}:\hat{\gamma}=\hat{D}\hat{\alpha}}{\operatorname{argmin}} ||\hat{\alpha}||_{\ell_1}. \tag{6.10}$$

The dictionary, $\hat{D}$, now becomes more skewed since the within-subject variation is minimised and between-subject variation is maximised. The more the dictionary is skewed, the sparser the solution becomes (See Figure 6.2). This can avoid the misrepresentation of sparse signals when there is similar global effect on the raw feature vector, and it results in a more robust sparsifying solution for the ℓ_1 norm minimisation. Figure 6.3 shows the comparison of sparse solutions resulting from an example test subject (subject 25) on the labelled dictionary (35 subjects, each of them having five feature vectors). Each bar shows the coefficients of the test subject on the dictionary. Using MDA, coefficients for the associated subject are significantly higher compared to the others and the coefficients are more sparse, compared to a PCA only approach.

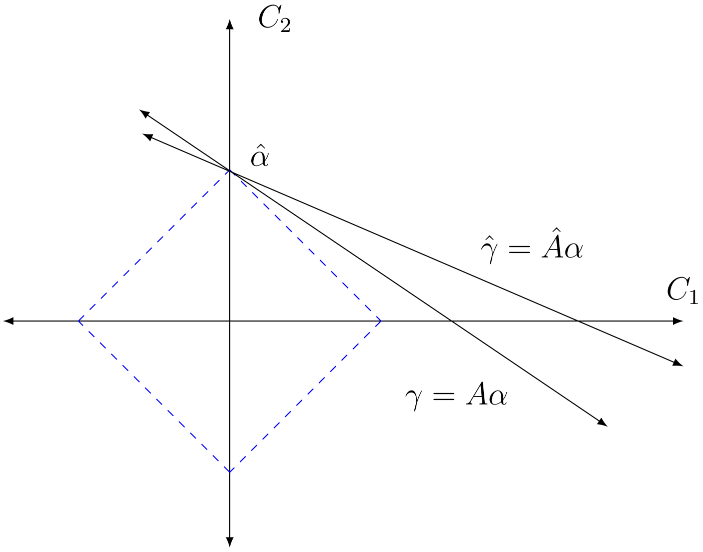


Figure 6.2: Influence of MDA input space on sparsest solution. Objective function using the MDA input feature space becomes more skew towards to the class labels that results more chance to meet the L1 ball in corresponding class.

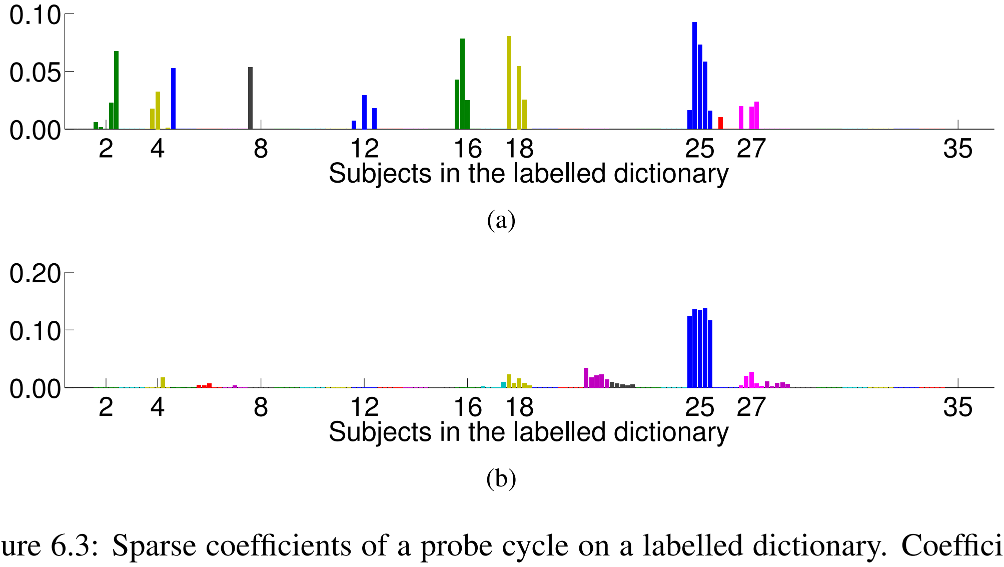


Figure 6.3: Sparse coefficients of a probe cycle on a labelled dictionary. Coefficients for the subject 25 are computed using l_1-norm optimiser on (a) component feature space and (b) discriminant feature space using MDA.

6.2.5 Classification

The sparse reconstruction solution via ℓ_1 norm produces sparse coefficients for a given test subject, γ, with respect to the training subjects in the dictionary. As explained in Section 6.2.1, ideally the non-zero elements in the sparse coefficients should represent the test subject. However, due to complicating factors and modelling errors, the sparsest solution can produce non-zero entries that correspond to multiple subjects. However, the most significant entry can be found by individually analysing the strength of coefficients associated with multiple subjects, by reproducing and comparing the y for each subject.

To do this, a characteristic function that selects the coefficients associated with the corresponding subject is defined as follows,

$$\delta_i(\alpha) = [0, ..., 0, \alpha_{i1}, ..., \alpha_{ik}, 0, ...0], \tag{6.11}$$

where $\delta_i(\alpha)$ represents only the coefficients corresponding to i^{th} subject from α. Now, a test subject γ can be approximately reconstructed from contributing coefficients of each subjects in the gallery as $\gamma_i = D\delta_i(\alpha)$ where γ_i is the approximation of γ respect to the ith subject. These approximations are used to compute the residual distance (RD between the actual γ and the approximated γ for each subject and the test subject is assigned to the subject that scores the minimum distance as follows,

$$\overset{\min}{i}\ RD_i(\gamma) = \left\|\gamma - D\delta_i(\alpha)\right\|. \tag{6.12}$$

6.2.6 Experiments and Results

To compare the effectiveness of the extended-SRC-based classifier for the gait recognition, the best performing truncated GEI at the height of 0.4 is computed as explained in Chapter 3 on the CASIA database. These GEI features are wrapped and transformed into the discriminative domain using the learned PCA and MDA basis on the training data. Both NN and SRC-based classifiers are used to compute the similarity scores and ROC curves and rank scores are used to compare the recognition performance.

Figure 6.4 compares the verification and identification performance of the SRC-based classifier with its nearest neighbour. Both results show improvement when SRC is used, however, ROC curves are significantly boosted, particularly in inter-class test-cases. This shows the robustness of the SRC-based classifier on the discriminated gait feature domain.

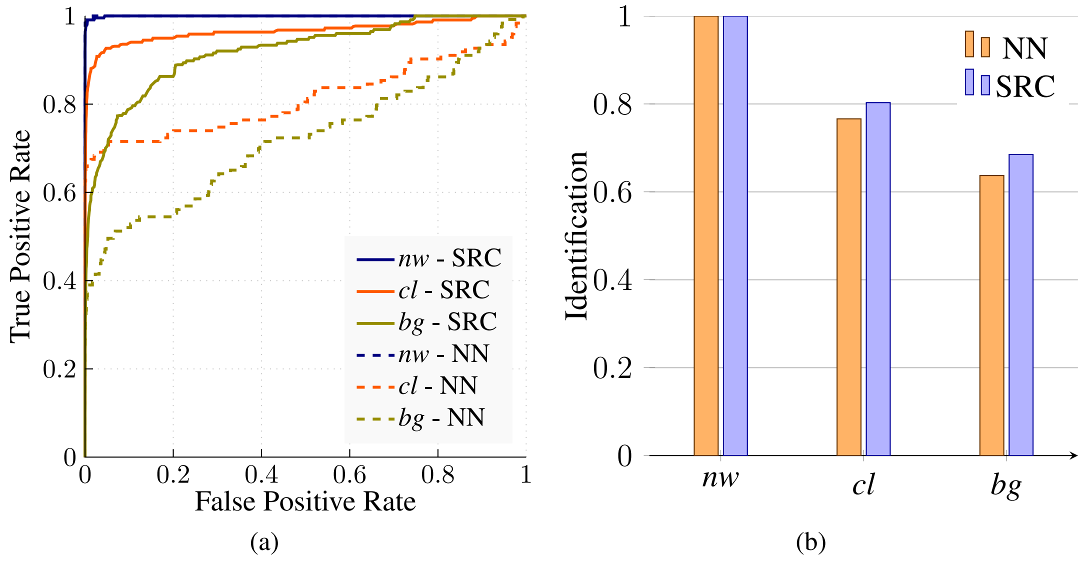


Figure 6.4: Recognition performance compares SRC and NN on the CASIA database ((a) ROC curves and (b) rank 1 cumulative match scores).

6.3 Local Histogram Feature Descriptors

This section explains the solution proposed for the issue of improper registration in the inter-class test cases. Due to the nature of the GEI, various forms of errors can be introduced into the feature image. Firstly, errors in the segmentation, as well as actual appearance changes in the subject, can cause changes to the appearance of the GEI. Poor normalisation and alignment of individual silhouettes can also contribute to artefacts, effectively causing a blurring effect on the final GEI.

Finally, the registration between the probe and gallery GEIs can also be an issue when performing classification. This can be addressed by using local grid-based histogram methods, as its coarse spatial sampling introduces some tolerance to these alignment issues.

Though more importantly, the use of histograms allows linear comparisons of feature values that are not scalar in nature, such as those produced from local binary pattern (LBP) and local directional pattern (LDP). LBP features were originally used as a texture descriptor, though have since been successfully applied in face recognition for their robustness to appearance changes due to illumination and pose, issues that are similar to those described above for GEIs. The LDP is an extension to the LBP, providing superior performance.

Another commonly used local histogram feature descriptor is the HOG. HOG has been used in gait recognition [59], but has yet to be applied directly to GEIs. The effectiveness of HOG and LDP in gait recognition context are explored by applying them to the GEIs.

A new feature descriptor has been proposed, which we call the histogram of weighted local directions (HWLD), borrowing concepts from both HOG and LDP. Like the HOG, the proposed method is a histogram based on local gradient directions. However, the raw directions are discrete, mapped to each pixel's eight neighbours, with the magnitudes in the possible directions determined from directional response kernels commonly used in LDP. Weights are applied based on the relative strengths of the directions at each pixel, like the LDP. This method takes the strengths of the two base systems, in that it keeps the richer local appearance description offered by the LDP, yet keeps a more compact feature size possible with the HOG. The second point allows us to extend the HWLD to 3D volumes (will be explained in Section 6.3.4 in Chapter 5), avoiding the prohibitively large number of histogram bins required for a similar LDP implementation.

6.3.1 Local Directional Pattern

The local directional pattern is an extension of the local binary pattern, and has been shown to be a robust feature for use in various face recognition applications [161]. Like the LBP, it assigns an eight bit binary key to each pixel of an input image (in this case, a GEI), representing the local appearance at that pixel.

The effectiveness of the LDP for gait has been explored, by extracting BCLDP features from GEI templates. The computed binary-coded image represents both the temporal movements of each pixel in local directions and the spatial position, and hence it better represents the gait of a person. Since much of the static appearance is removed by this feature, it is stable to many of the challenges present in appearance changes.

To compute the LDP, a set of eight directional kernels are applied to extract the dynamic response in each of the eight neighbouring pixel directions. The kernels used are based on the Kirsch compass kernels [161], that extract the directional dynamic response of every pixel in the GEI image. The directional Kernel, K_i with a dimen-

$$\underset{NorthEast(F_1)}{\begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & -3\end{bmatrix}} \quad \underset{East(F_2)}{\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}} \quad \underset{SouthEast(F_3)}{\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}} \quad \underset{South(F_4)}{\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}}$$

$$\underset{SouthWest(F_5)}{\begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}} \quad \underset{West(F_6)}{\begin{bmatrix}5 & -3 & -3\\ 5 & 0 & -3\\ 5 & -3 & -3\end{bmatrix}} \quad \underset{NorthWest(F_7)}{\begin{bmatrix}5 & 5 & -3\\ 5 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}} \quad \underset{North(F_8)}{\begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}}$$

Figure 6.5: 2D local directional kernels.

sionality of $m \times m$ in i^{th} direction is defined as follows,

$$K_i(r,c) = \left\{ \begin{array}{cc} (m^2 - |N_i|) & (r,c) \in N_i \\ 0 & r = c = \frac{m+1}{2} \\ -|N_i| & otherwise \end{array} \right\}. \tag{6.13}$$

Where the directional nearest neighbour element, N_i satisfies the following nearest neighbour condition,

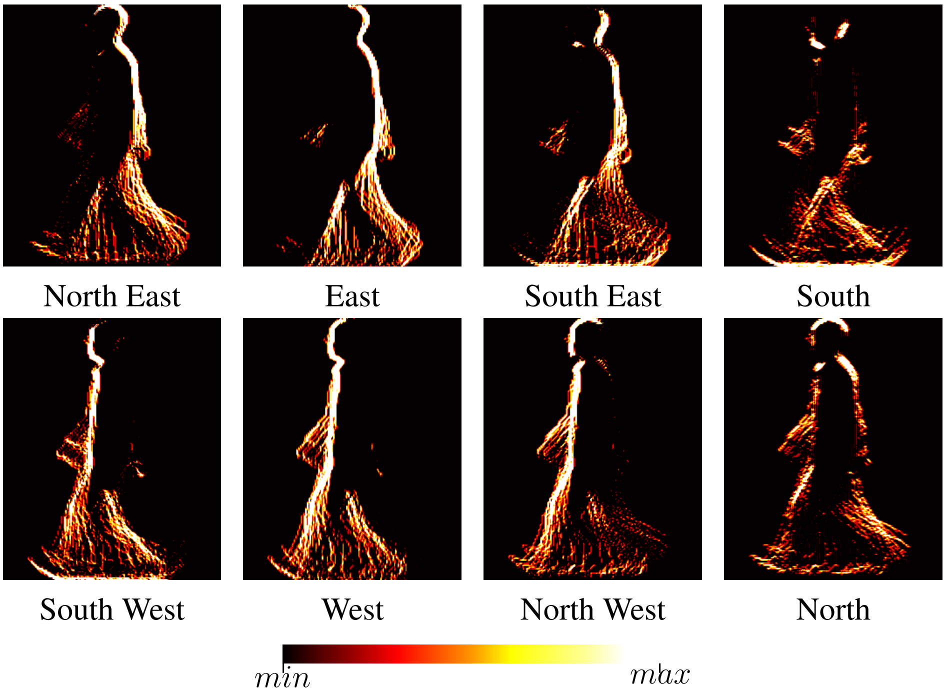
$$N_{dist} = \sqrt{\frac{m-1}{2} \times (D-1)}, \tag{6.14}$$

where N_{dist} is the directional distance between N_i and the directional element, C_i, and D is the template image spatial dimensionality included to generalise for multi-dimensional kernels. The smallest kernel that accommodates all the nearest neighbouring directions can be defined by selecting $m = 3$ as explained in [61]. For 2D, by choosing $m = 3$, there are eight unique directional kernels to accommodate all the neighbouring directions of each pixel (Figure 6.5).

To represent the temporal dynamics in each direction, the GEI is convolved with the respective directional kernel and this produces the dynamic response values ($k_0, k_1, ...,$-k_7) for each pixel in the GEI. The resultant dynamic response images (DRI) represent the dynamic characteristics of the GEI in each direction as shown in Figure 6.6.

The dynamic response values are sorted in descending magnitude, and the top n values are selected. The eight possible directions each correspond to a bit in an 8-bit value. The selected n bits are set to 1, and the resulting value is the feature ‘key’ for that pixel. For the purposes of this paper $n = 3$.

The resulting feature map is partitioned into a grid, and a histogram is computed for each local patch. 56 histogram bins are used, corresponding to the 56 unique key



North East

East

South East

South

South West

West

North West

North

Figure 6.6: Directional response images of an example GEI.

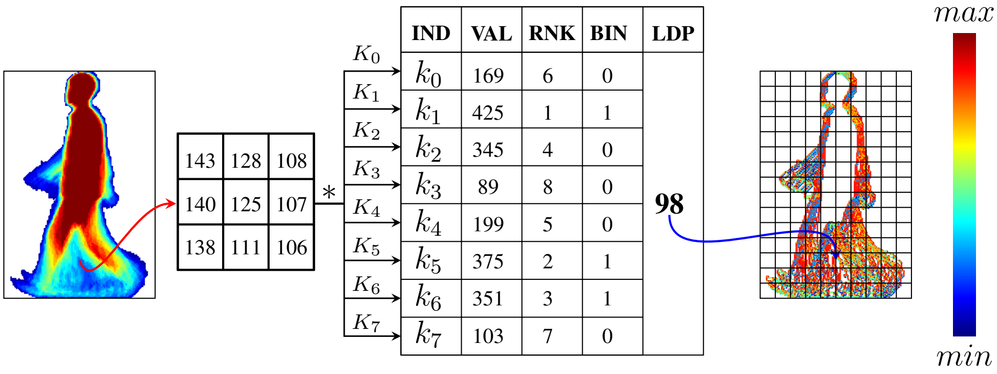
values ($^8C_3$). The histogram values from each region are then concatenated to form the final feature vector. Figure 6.7 shows the process of computing the LDP feature from a GEI.

LDP however, is not suitable for applications using 3D data, such as the GEV. With 26 neighbours per voxel, the number of unique keys, and therefore required histogram bins, become infeasible. At $n = 3$, the number of bins required is 2600. Therefore, it is required to optimised the LDP feature vector in the sense that optimises the dimensionally without loss of gait information to extend it to 3D.

6.3.2 Histogram of Oriented Gradients

The histogram of oriented gradients (HOG) is a very common feature descriptor used for object detection and recognition. Though it has been used in face recognition [151], it is less popular than LBPs due to poorer performance in that context. Recently, it has also been applied to gait recognition [59], showing significant improvements over a GEI/PCA baseline [98].

In [59], the HOG is applied directly to the raw image, and the magnitudes of the gradients are weighted by the silhouette mask. The HOG features are then averaged over the gait cycle to arrive at the final feature vector. In our analysis, the HOG operator



143	128	108
140	125	107
138	111	106

IND	VAL	RNK	BIN	LDP
k_0	169	6	0	98
k_1	425	1	1	
k_2	345	4	0	
k_3	89	8	0	
k_4	199	5	0	
k_5	375	2	1	
k_6	351	3	1	
k_7	103	7	0	

Figure 6.7: Computation of binary-coded LDP from a GEI image. The GEI is convolved with the directional kernels. The resulting values are sorted by rank (RNK), and the top three are assigned a 1 in its corresponding bit in the binary number (BIN). The decimal representation of this number is the LDP value for the pixel. The GEI is partitioned into a grid and a histogram is computed for each patch from the LDP values.

is applied directly to the GEI image instead. The GEI encodes the temporal features into the image appearance; applying the HOG to individual frames and then averaging will lose this information.

To compute the HOG feature, the gradient vector at each pixel in the GEI is first found. The 1st order gradient operators ($[-1, 0, 1]$ and $[-1, 0, 1]^{\mathrm{T}}$) are applied to extract the horizontal (g_h) and vertical gradient (g_v) magnitudes, which are then combined to calculate the final gradient magnitude (g) and orientation (θ) as below,

$$g = \sqrt{{g_h}^2 + {g_v}^2} \qquad \theta = atan2(g_h, g_v). \tag{6.15}$$

The image is then partitioned into a grid, with the gradients placed into histograms at each patch based on the gradient orientation. The histograms contain nine equal-width bins, with each gradient weighted by the gradient's magnitude.

Applying the HOG to the GEV is possible, as the number of bins used can be adjusted. Selecting the values required to represent an angle in 3D space can be cumbersome, with the directional kernels used to compute the LDP providing a more elegant solution. This motivates the development of the histogram of weighted local directions.

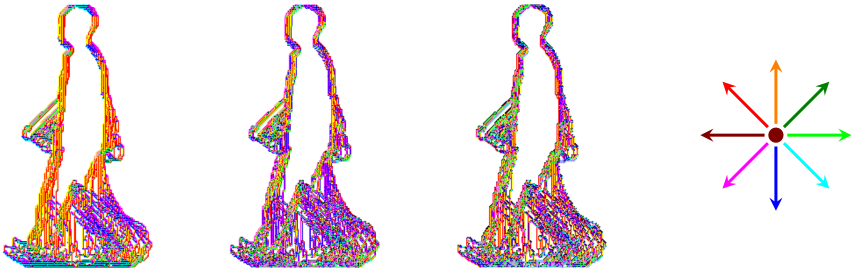


Figure 6.8: First three significant directional dynamic responses of a GEI. The colour code represents the response direction.

6.3.3 Histogram of Weighted Local Directions

The initial motivation for developing this new feature is to extend the LDP into 3D for use in frontal gait energy volumes, introduced in Chapter 5. Due to the binary pattern coding used however, the number of unique code combinations (and therefore, the required number of histogram bins) increases to 2600 with the 26 possible directions in 3D voxel space and $n = 3$. This number rises rapidly as n increases, with 65780 unique values at $n = 5$.

The solution is to simply remove the binary pattern coding from the algorithm, moving towards more HOG-like implementation. However, the proposed algorithm retains the eight discrete directions, with magnitudes computed using the directional kernels. Multiple gradient directions for each pixel are still used, and this should provide a richer description of the local appearance than simply using one, as many pixels would not have a dominant gradient direction. The contributions to the histogram are weighted, as a function of its directional response values. The proposed method is termed as the histogram of weighted local gradients (HWLD).

The initial computation of the HWLD is identical to that of the LDP; the directional kernels are convolved with the GEI to extract the directional response values, which are used to rank the directions in terms of their magnitude. As an example, the first three directional images are shown in Figure 6.8. The GEI is partitioned into a grid, and a histogram is formed for each patch. The first n directions at each pixel are used to populate the corresponding histogram, though their contribution is weighted by their magnitudes. A higher value of n is more sensitive to appearance changes. Based on the cross validation evaluation, n is set to 3 for optimal performance. The weighting factor, w, is proportional to the relative dominance of each direction at the pixel as follows,

$$w_j = \|k_j\| \left(\sum_{i=0}^{7} \|k_i\| \right)^{-1}, \tag{6.16}$$

where j is a given direction index, and k is the response value.

6.3.4 Applying HWLD to GEV

Since reconstruction of frontal volumes suffers from missing regions in the depth image (similar to segmentation issues encountered when generating silhouettes), the voxel volumes generated are coarse. This causes performance degradation for gait recognition, primarily when applying appearance based methods such as GEV. The comparative performance on ROC curves of frontal carrying goods also suggests that, though frontal GEV performs better, it can be affected by heavy appearance changes.

To apply HWLD to 3D, local directional kernels for 3D image templates are defined based on the Kirsch compass kernel as explained in Section 6.3.1. There are twenty-six $3 \times 3 \times 3$ dimensional kernels defined for each neighbouring directions. Like the Kirsh compass kernel, the direction facing voxel and its direct neighbours are given positive values, while all others, excluding the centre voxel, are assigned negative values. These positive and negative values satisfy the kernel properties as below,

$$\begin{gathered} \sum_{r=1}^{m} \sum_{c=1}^{m} \sum_{d=1}^{m} K(r,c,d) = 0, \\ \|K\| = 1. \end{gathered} \tag{6.17}$$

Based on this, the values of the i^{th} directional kernel can be computed as follows,

$$K_i(r,c,d) = \left\{ \begin{array}{cc} (m^3 - |N_i|) & (r,c,d) \in N_i \\ 0 & r = c = d = \frac{l+1}{2} \\ -|N_i| & otherwise \end{array} \right\}. \tag{6.18}$$

Since the number of positively and negatively weighted elements are different, to satisfy the equality conditions, the kernel shown in Equation 6.17 is redefined as follows,

$$\hat{K_i} = \frac{K_i}{\|K_i\|}, \tag{6.19}$$

where $i = \{1, 2, ..., m\}$.

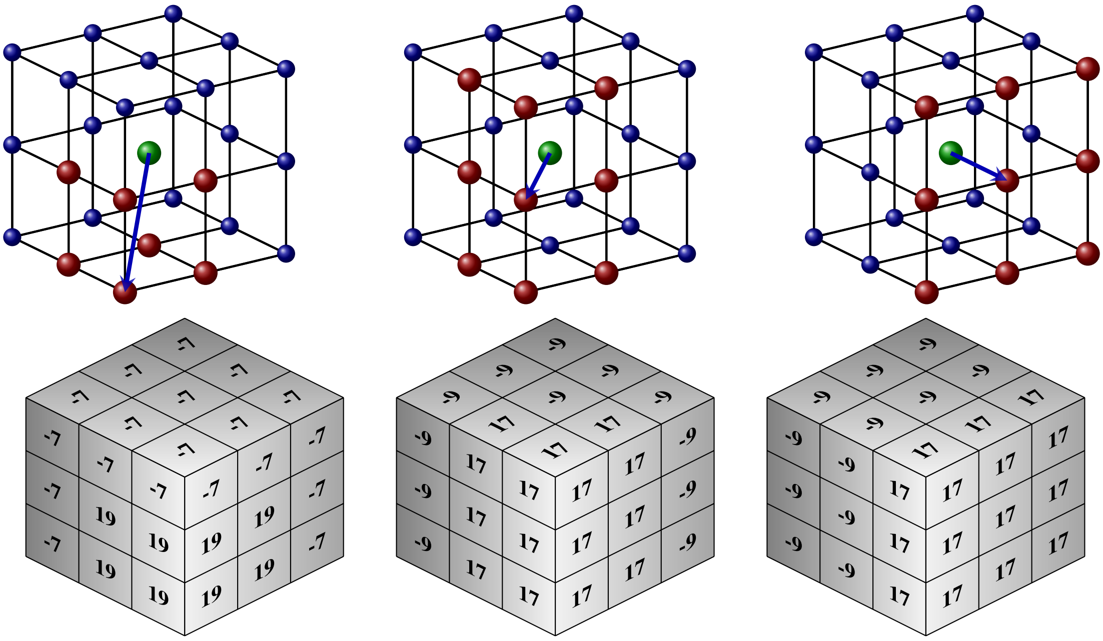


Figure 6.9: 3D local directional kernels. Three unique cases are presented, showing the kernel direction and the corresponding weight values. The other 23 kernels can be obtained by applying rotational transforms to these.

Similar to the 2D approach (see Section 6.3.3, by choosing the immediate neighbourhood voxel directions, there are 26 kernels with a dimension of $3 \times 3 \times 3$ that can be defined to extract local directional gait features. Example directional kernels in 3D are shown in Figure 6.9. The upper images show the directional patterns of the kernel and the lower images show the assigned values based on Equations 6.18.

Significant directional volumes (SDVs) are computed using these 3D kernels, based on a method similar to that in Section 6.3.3, where GEVs are used as an analogue to GEIs. Following that, the histogram feature for each 3D patch of each SDV is computed and by adding the weighted histograms of the first n SDVs and the WLDP feature, w, is produced. For optimal performance, n is chosen as 5 for the 3D case. Note that a similar BCLDP requires a 65780 bin histogram per patch, compared to the 26 bin histogram for the WLDP. Example cross sectional views of the first five SDVs are shown in Figure 6.10.

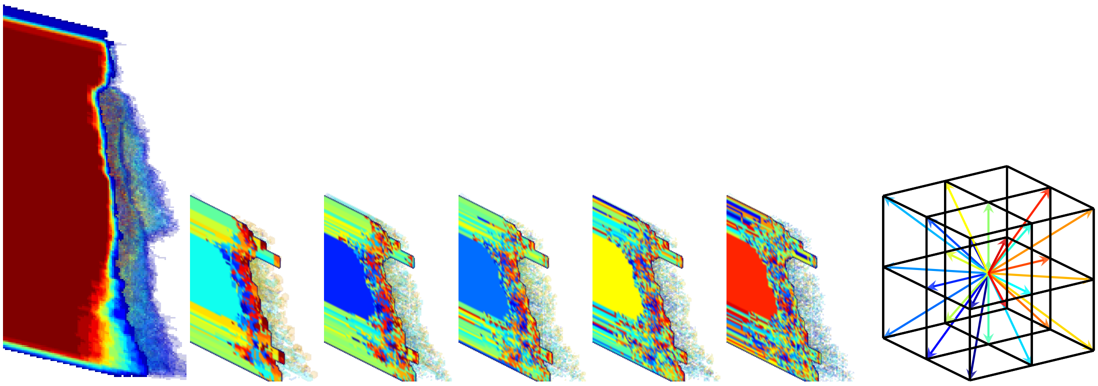


Figure 6.10: Significant directional volumes of an example GEV. The cross-sectional view of a GEV and the corresponding first five significant directional volumes of the lower body only.

6.3.5 Experiments and Results

HWLD evaluation in 2D

Profile views from the CASIA dataset B [141] and the OULP database [46] are used to evaluate the proposed and baseline LDP approaches in 2D. Both identification and verification experiments are performed, with CMS and ROC results compared. Intra and inter-class evaluations are performed using the framework outlined in the respective datasets as explained in Section 3.2 in Chapter 3.

LDP, HOG and HWLD features are computed from the GEI as explained in Section 6.3 using a patch size of 5×5. PCA and MDA is applied to the features, with classification performed using SRC as explained in 6.2.

Two sets of silhouettes (original-S1 & cleaned-S2) from the CASIA database (explained in Chapter 3) are used to evaluate the effectiveness of the proposed HWLD features in the presence of segmentation noise.

From the ROC curves in Figure 6.11, it can be seen that the proposed HWLD is able to perform equally, or better, than the other techniques in all test cases and achieves the state-of-the-art performance on the CASIA dataset. Also, observed were good performances of the LDP and HOG, with LDP favouring the *nwbg* case and HOG favouring *nwcl*. This result is slightly unexpected as LBP/LDP approaches generally provide better performances than HOG in face recognition. This could simply be due to the different nature of a GEI compared to a face image, though the performance of the LDP may have been limited due to the small patch size, as fifty six histogram bins are being populated with twenty five samples, compared to eight or nine bins in

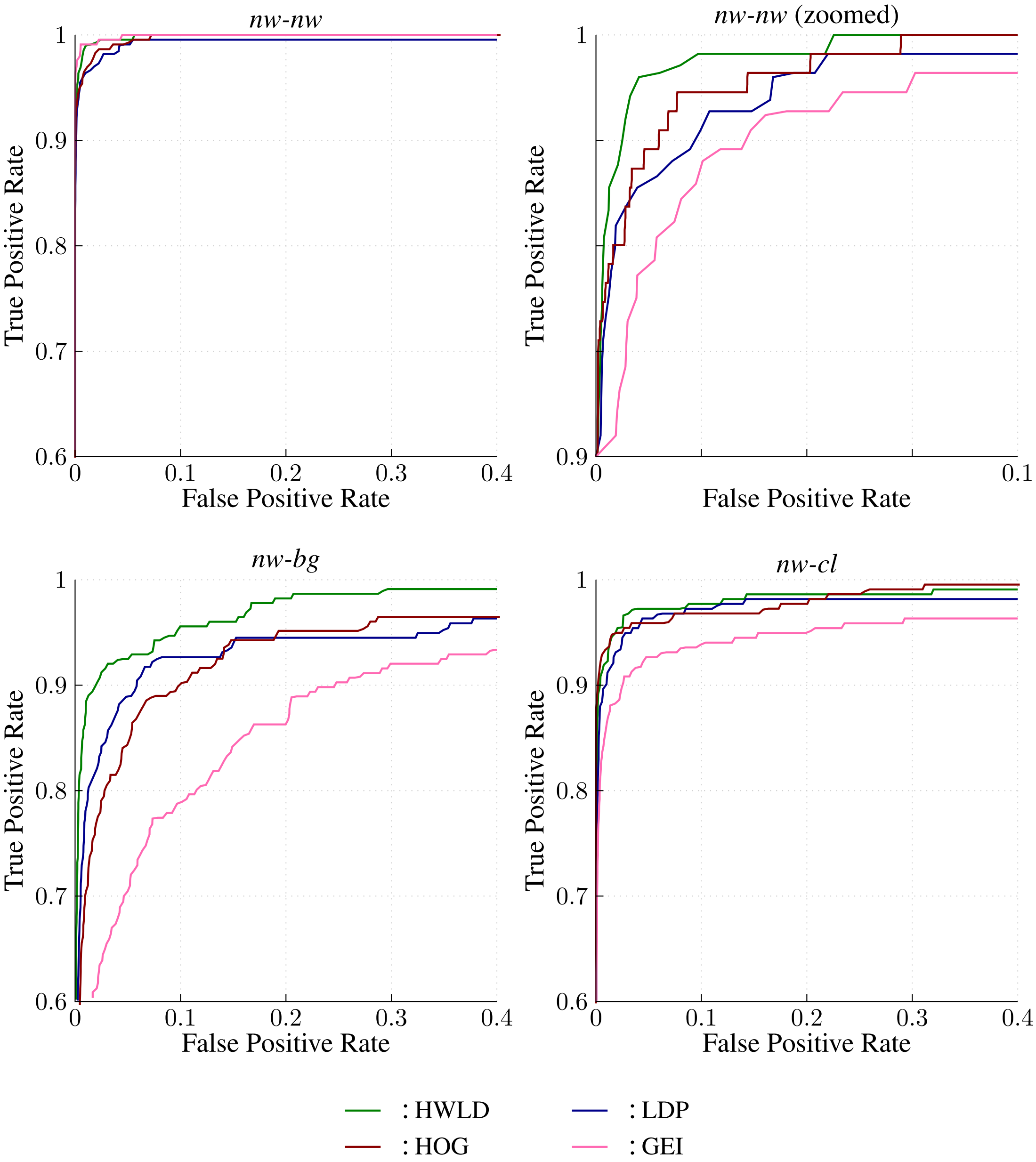


Figure 6.11: Comparison of patch-based descriptors.
ROC curves of patch based descriptors including the proposed HWLD, comparing with the baseline GEI with PCAMDA and the (**Exp 1**).

the HWLD and HOG [161] finds slightly improved performance with a patch size of 10×10 compared to 5×5 (See Appendix A).

In the Chapter 3, it has been shown that a traditional GEI performed poorly on the silhouettes distributed with the CASIA database as the poor segmentation kills the registration, particularly in inter-class test cases. One of the motivations for the proposed HWLD method is to provide tolerance to these registration issues through its coarse spatial sampling. This functionality has been evaluated by comparing the stability of the recognition performance on both improved segmented silhouettes (S2) and the poor segmented silhouettes (S1). In Figure 6.12, only a small variation in recognition performance is observed using the local patch methods particularly with the proposed HWLD (Figure 6.12). Contrasting this, with the greater performance variations obtained when using just the GEI, demonstrates the greater tolerance to segmentation errors in these algorithms.

It is also noted that better performance is achieved in the patch-based methods when PCA and MDA are not applied to the feature vectors in inter-class tests. The reason for this is unknown, though it has been speculated that this could be due to 'over-fitting', with the SRC dictionary templates failing to properly generalise to variations in the subjects' appearance.

The proposed approach is also evaluated on the OULP dataset [46]. Only intra-class test cases are considered by following the testing protocol in [46], as explained in Section 3.2.3 in Chapter 3. The first sequence is in each group is used as the gallery and the other one is used as probe.

Table 6.1 shows the cumulative match scores for Rank $= 1$ and Rank $= 5$ and AUC metrics to compare the recognition performance of proposed descriptors. Again it shows the effectiveness of the proposed HWLD algorithm with the AUC close to 1 for the A-All class. This provides further support for the over-fitting hypothesis previously mentioned, as these are intra-class tests.

The results for the GEI are lower than the GEI implementation in [46]. The reason for this is likely due to using only the lower half of the GEI. Since this is an intra-class evaluation, there are no significant appearance changes in the upper body for us to ignore.

HWLD evaluation in 3D

Experiments in 3D are conducted using the DGD database proposed in Section 5.2.1 in Chapter 5. Following the protocols defined with the database, normal-walk sequences

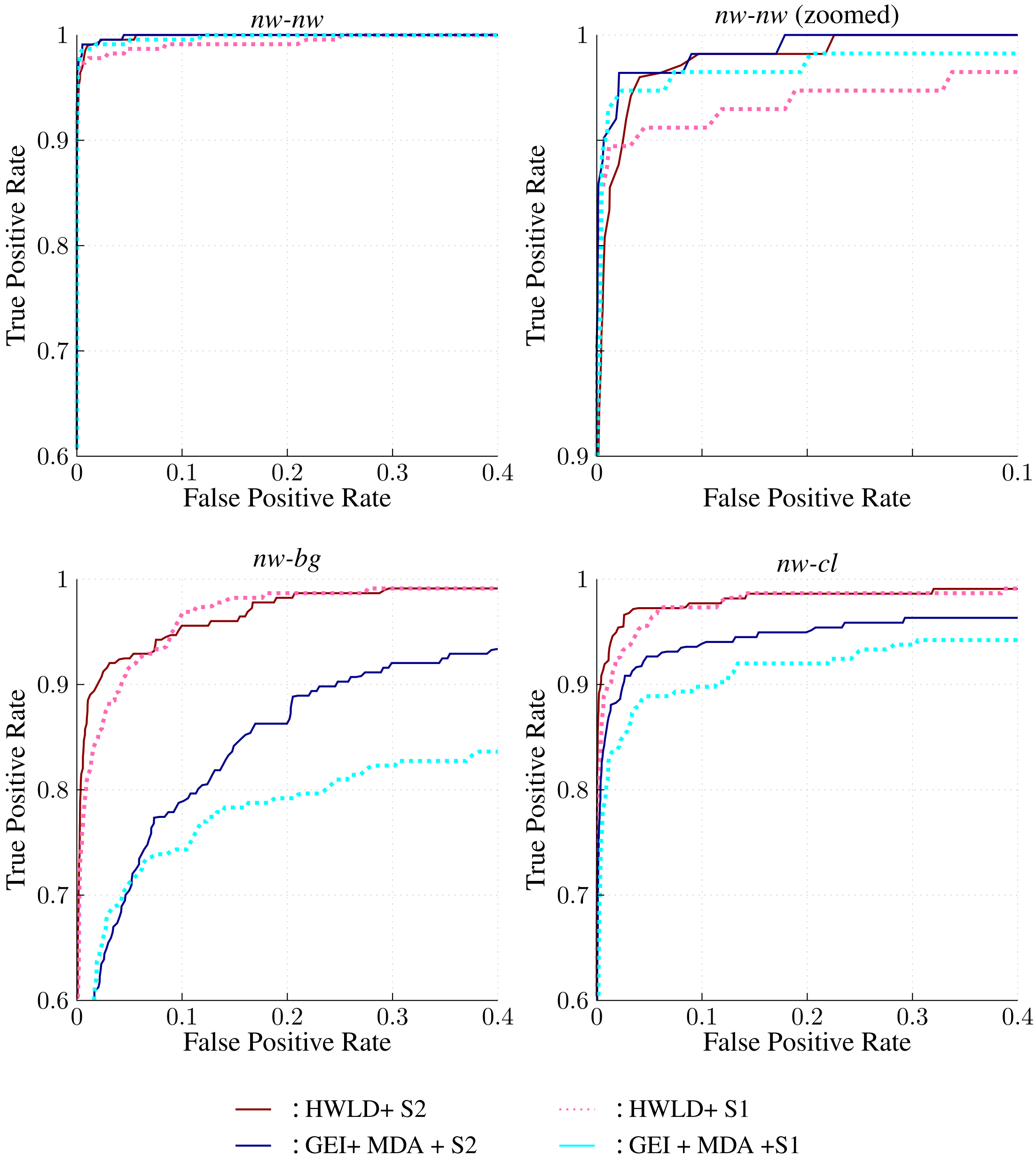


Figure 6.12: Comparison of HWLD and GEI+MDA approaches on the presence of noisy segmentation (**Exp 2**).

Feature	A-85			A-All		
	Rank 1	*Rank 5*	*AUC*	*Rank 1*	*Rank 5*	*AUC*
HWLD	87.7	94.7	0.992	95.5-	98.5	0.999
LDP	83.7	91.2	0.992	95.1	98.3	0.999
HOG	56.5	74.5	0.954	75.9	89.9	0.984
GEI+ PCA	81.1	92.0	0.985	90.8	97.2	0.997
[46]	85.7	93.1	-	94.2	97.1	-

Table 6.1: Recognition Performance on OULP-C1V1-A.

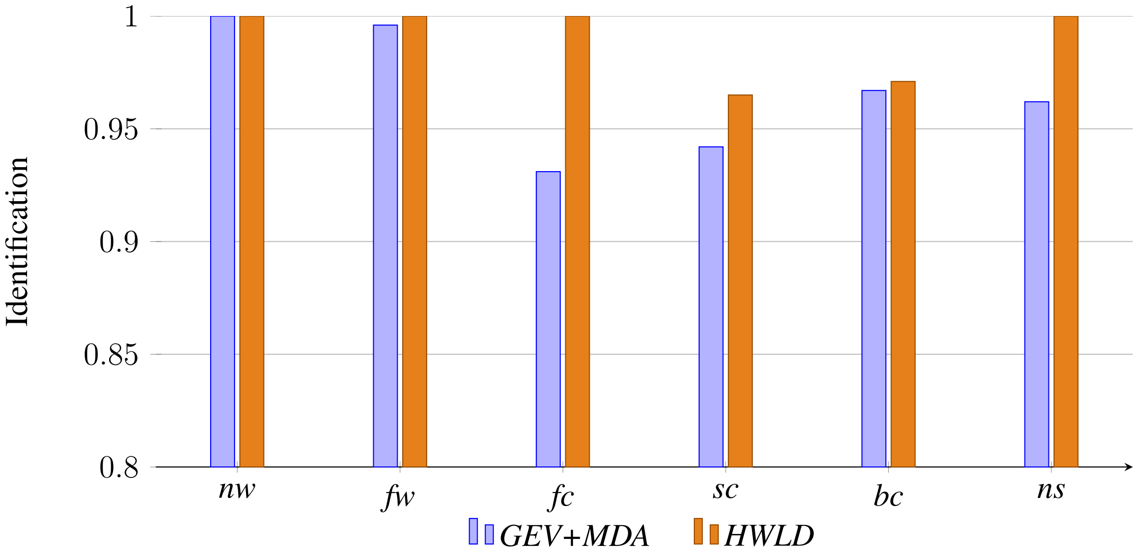


Figure 6.13: HWLD evaluation in 3D using the DGD database. Rank-1 cumulative scores are compared.

are used for intra-class and in inter-class tests, all five nw cycles are used as the gallery, and all available cycles in each of the other classes (fw, sc, fc, bc, ns) are treated as the probe in their respective experiments.

GEVs from frontal depth images in the DGD are extracted using the method described in Section 5.2. Similar to the 2D experiments, HWLD features are extracted as explained in Section 5.2. The Rank-1 cumulative match scores for these experiments is presented in Figure 6.13. A slight, but consistent improvement has been observed over the baseline FGEV implementation.

6.4 Summary

In this chapter, sparse representation-based classification for gait recognition has been explored and the SRC-based classifier on the discriminated domain features is proposed as the optimised solution, particularly when the training data does not include all variations.

A novel histogram descriptor, the Histogram of Weighted Local Directions (HWLD) is also has been proposed for use in appearance-based gait recognition. The proposed HWLD features demonstrate state-of-the-art performances, showing 15-20% improvements in recognition identification score at rank 1 in inter-class tests, over existing implementations on the CASIA dataset. Superior performance of the feature on the high population OULP dataset, which contains more than 3000 subjects, shows that the proposed method is stable over a large population. The HWLD can also be easily extended to 3D, with evaluations using GEVs on the DGD dataset beating all known results.

Furthermore, the local histogram feature extraction techniques have been demonstrated as much more stable to minor segmentation errors. They also show improved performance in the absence of feature conditioning processes such as PCA and/or MDA in inter-class tests.

Chapter 7

Back-filled Gait Energy Image

7.1 Introduction

Gait energy-based features are popular due to their high recognition rate and simple approach [101, 117]. However, like many other appearance based features, better recognition performance is achieved when video is recorded from the side, as the gait features, particularly the motion of the legs, are better captured in that view. Limited gait recognition research has been performed on a frontal perspective [44], due to its inability to capture the gait dynamics as the walking direction is occluded in a 2D frontal image. The use of stereo cameras [162] and other depth-based sensors (*e.g.* Microsoft Kinect) overcome this issue by working in 3D. In Chapter 5, state-of-the-art recognition results are achieved on the frontal view on the CMU MoBo database using the GEV features from synthesised frontal depth images.

A frontal perspective has various advantages over that of the side, such as use in narrow corridors, where the limited field-of-view of cameras may prevent the recording of a complete gait cycle from the side. They can also be easily integrated into biometric portals such as that used in the Multiple Biometric Grand Challenge (MBGC) [163]. However, there are also situations where gait recognition from a side-view is preferable, such as in surveillance where the distances involved may be unsuitable for many depth-sensing devices, or where depth-sensing hardware may simply not be present. These two different capture modalities operate in differing image domains, with the gait features used in existing approaches specific to each. This prevents sharing of information without the use of view transformation models.

As a part of the research on frontal GEV, experiments have been evaluated to explore the loss of recognition performance due to the discarding of back leg information.

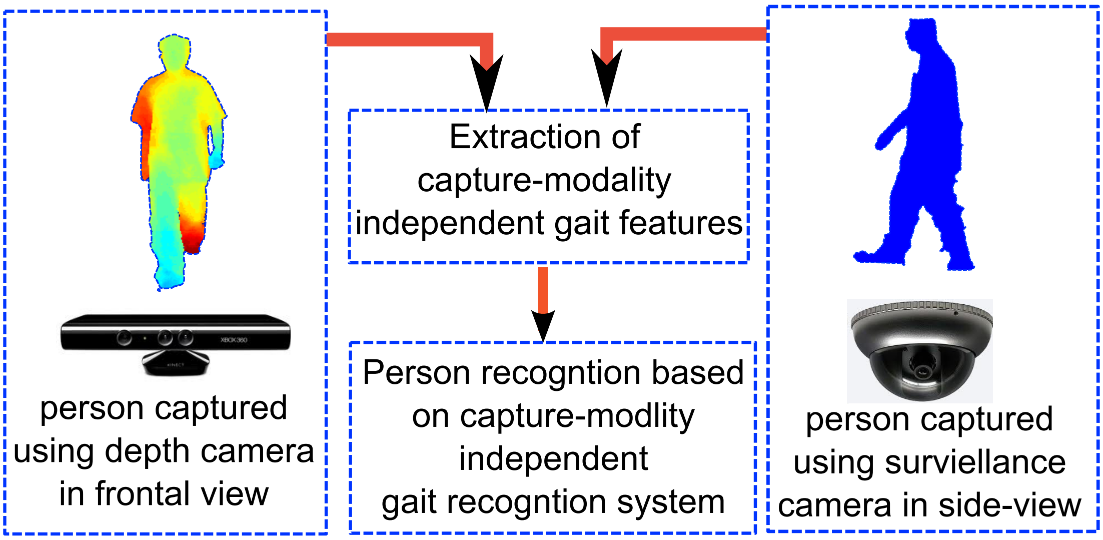

person captured using depth camera in frontal view
Extraction of capture-modality independent gait features
Person recogntion based on capture-modlity independent gait recogntion system
person captured using surviellance camera in side-view

Figure 7.1: An example capture modality independent gait recognition system.

In those experiments (see Section 5.3 in Chapter 4), a small improvement is achieved using the GEV on synthesised frontal 3D volume compared to the GEV from full 3D volumes. This improvement could be attributed to the suppression of noise and appearance changes unrelated to the underlying gait kinematics found in the back surface of the model. Can such a method be applied to a 2D silhouette? Back-filling from the frontal contour, a side view silhouette will result in the removal of the back leg, leading to a loss of gait information, but is this information needed to accurately perform gait recognition? These these questions has been answered in this Chapter by developing such a gait energy feature, which is termed the BGEI, or backfilled gait energy image.

The proposed backfilled gait energy feature can be constructed from both side view silhouettes and frontal depth images. This allows the feature to be applied across differing capturing systems using the same enrolled database, such as in a system using both frontal depth cameras mounted on biometric portals and general surveillance cameras as shown in Figure 7.1. The effectiveness of this proposed framework is explored by experimentally demonstrating how the BGEI can be used to match subjects across the two modes. This is performed on an expanded DGD database which contains 37 subjects under various walking conditions captured from the front using a depth camera. Eight of these subjects also have gait sequences recorded from the side in order for us to perform the cross-modality experiments.

Using this database, the BGEI has also been evaluated against the GEV in intra-capture modality experiments, comparing the BGEI to domain-specific features. A similar experiment was performed with the GEI using the CASIA dataset B.

To perform the recognition in our experiments, sparse representation based classification (SRC) is used as explained in Chapter 6.

The proposed feature, BGEI represents gait dynamic based on frontal contour. Frontal contour is less susceptible for perspective distortion for view angle changes compare to full silhouette. Based on that, BGEI also has been evaluated for view-independency and better performance is achieved.

The remainder of this Chapter is organised as follows. Section 7.2 outlines the baseline feature extraction methods used to benchmark the proposed algorithm while Section 7.3 illustrates the extraction of the proposed backfilled gait energy image (BGEI). Experiments and results for BGEI evaluations for cross capture modality are shown in Section 7.4. Section 7.5 explains the applicability of BGEI in view invariant gait recognition, followed by the conclusion in Section 7.6.

7.2 Gait Energy Features

To achieve a cross-capture-modality gait recognition system, the gait features that provide best performance in each specific domain are considered. In this thesis, the popular and high performing algorithms in side-view and frontal-view gait recognition are considered.

As discussed in Chapter 3, GEI is a simple and effective gait feature that performs comparatively well on the side-view [164]. However, its performance is dominant only in the side-view. Therefore GEI is chosen as a domain-specific feature in side-view to evaluate the proposed algorithms.

Traditional 2D frontal images are poorly suited for gait recognition due to the inability to capture dynamic details of gait [44]. However, depth images, either from stereo cameras or other depth sensing devices, can be used to alleviate this. In Chapter 4, the gait energy volume is proposed to exploit the robustness of gait energy features in the 3D domain using these depth images. It is a volumetric extension of the GEI, where binary voxel volumes are used as an analogue to the binary silhouette. Both full body volumes and frontal surface reconstructions have been used. In the context of frontal-view gait recognition, constructed frontal surface volumes are used to generate frontal GEV. A frontal (or possibly even back) perspective is ideal as it does not suffer from occlusions between the legs, and in theory, it should contain all the relevant dynamic gait information as the hidden surface should only contain relative structure information (*i.e.* thickness of limbs and torso).

The GEI and GEV use two different capturing systems and extract features that are heavily dependent on those capture-modalities. As their representations are incompatible, there is a requirement to maintain a separate recognition system for each feature.

A new feature, the Backfilled Gait Energy Image (BGEI), is proposed that captures the essential but common gait information from the above two models and enables cross-modality comparisons where the user can enrol in a side-view and be recognised in frontal-depth or vice-versa.

7.3 Backfilled Gait Energy Images

The backfilled gait energy image (BGEI) operates on a similar premise to the frontal GEV, where the frontal surface of a model should contain all the relevant gait information. It takes only the frontal contour of the silhouettes and assumes that it contains sufficient information to perform gait recognition. By doing so, there is a possibility of losing some of the gait information as the back leg is no longer represented by this feature. This information, however, could potentially be unnecessary for, or at least contribute minimally, to the system's ability to discriminate between different people.

Since the frontal surface is available in both the side-view and frontal-depth, by applying the above concept, the BGEI becomes as a common feature for both.

For the side-view silhouettes, the BGEI is constructed by first back filling the binary silhouettes. For this, the front-most pixel on each row is found and from it, filled to the back of the image. These backfilled binary silhouettes are aligned based on the centroid of the frontal surface. The BGEI is then constructed from these silhouettes in the same manner as a GEI, by averaging within a gait cycle. Figure 7.2 shows example images of backfilled silhouettes and computed BGEI constructed from side-view images.

To create a BGEI from frontal depth images, first frontal binary voxel volumes are constructed as outlined in Section 5.2. These frontal volumes are projected into the sagittal plane to produce the backfilled binary silhouettes. These backfilled silhouettes are used to generate the BGEI following the similar method used in the side-view. Example backfilled silhouettes and the computed BGEI using depth images with corresponding GEVs are shown in Figure 7.3.

Computed features are transformed to the PCA domain, then to the MDA domain and similarity distances are computed using an src-based classifier as previously outlined in Section 3.3.3 in Chapter 3.

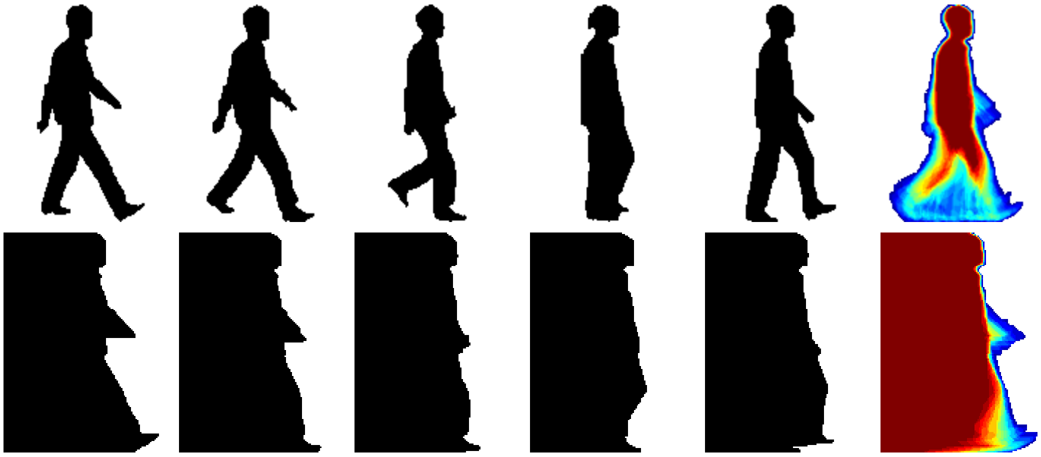


Figure 7.2: Computation of BGEI from side-view. Sample silhouettes of a gait cycle and computed GEI in top and corresponding backfilled silhouettes and computed BGEI in bottom.

7.4 Experiments and Results

7.4.1 Experiments

Two sets of experiments are carried out in this thesis to evaluate the BGEI on inter-capture modality and intra-capture modality platforms. Both the DGD as well as the CASIA dataset B [55] are used in these experiments.

To obtain our gait energy features, voxel volumes as well as silhouettes, need to be extracted from the databases. For the DGD, voxel volumes are constructed first by projecting the depth images into world coordinates. Segmentation planes are used to remove the background and a surface mesh of the subject is created in 3D. Holes in the data are interpolated, and the mesh is filled backwards to create the frontal binary volume. Gait cycles are identified based on detecting the oscillating pattern of the width profile in the volumes' lateral view (See Section 4.3.2 in Chapter 4). The GEVs and BGEIs for each gait cycle are then computed from these volumes as explained in Section 7.3.

For the side-view sequences in the multi-modal subset, silhouettes are extracted from the depth images as opposed to the colour images. This is chosen as the extracted silhouettes are of higher accuracy, and the poor lighting and sensor quality in the colour camera makes clean segmentation from the RGB images difficult. Note that the depth information of these side-view sequences is only used to obtain the silhouettes and not in the experiments themselves.

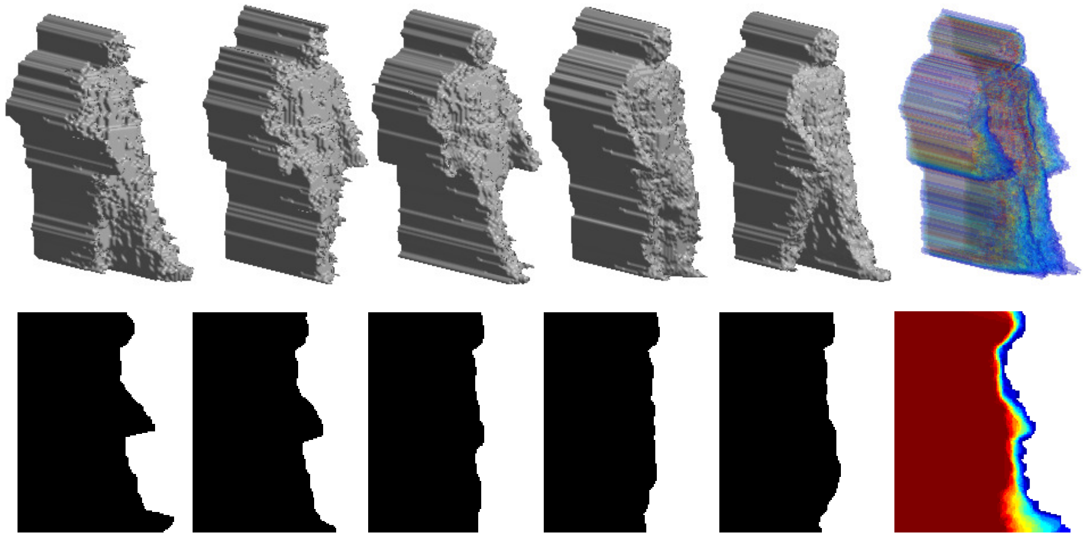


Figure 7.3: Computation of BGEI from frontal volume. Sample frontal voxel volumes of a gait cycle and computed frontal GEV in top and corresponding backfilled silhouettes and computed BGEI in bottom.

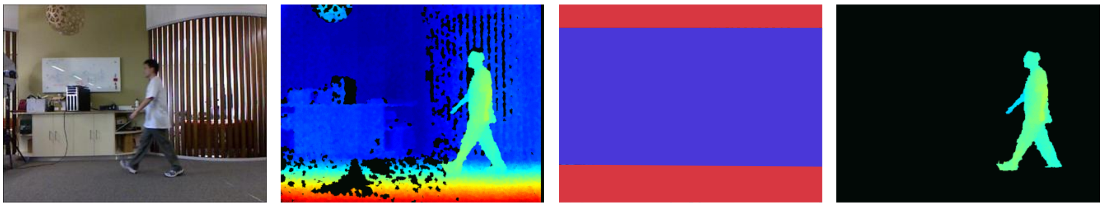


Figure 7.4: Segmentation of side-view images using plane segmentation on side-view depth image. Example side-view (from left to right) RGB image, depth image, segmentation planes and segmented depth image.

To do this, the DGD is extended with sequences captured from the side. Six subjects from the database are recorded under normal walking conditions at five sequences per each. Two further subjects are also recorded performing the five walk sequences from the side, as well as from frontal-depth. This brings a total of eight subjects as a multi-modal subset of the DGD database. Examples of this silhouette extraction are shown in Figure 7.4.

The gait cycles are detected based on the width profile, and the GEI and BGEI of the side-view sequences are computed according to Section 7.3. Only one gait cycle is extracted from each sequence in the DGD. For frontal depth sequences, the closest complete cycle is used in order to maximise the depth resolution (depth resolution in the Kinect sensor decreases with distance). For the side-view sequences, the central cycle in the image is used. This is to minimise any changes to the apparent size due to changes in distance as the subject moves across the camera's field of view.

In the CASIA database, dataset B, 90° (side-view) sequences are used. Background subtraction is used to extract the silhouettes from the video sequences. Graph cuts, similar to that used in [83] are used to improve segmentation quality. Once the silhouettes are obtained, gait cycle detection and gait energy feature construction is identical to the process used for the side-view DGD sequences. Two to three complete gait cycles exist in each side-view sequence, however, once again, only the centre-most cycle is used in the experiments.

All gait energy features are scaled to 96 pixel height. Only the lower half of the body is used to remove unwanted motion and appearance changes from the upper body. It also significantly reduces the computational cost by decreasing the feature dimension. This results in a height of 48 and width of 84 in the feature image. The GEVs use the same dimension in the sagittal plane, but with an additional depth of 60 voxels.

The gallery cycles in each experiment form the training set for our classifier. These are used to learn the PCA-MDA basis (T_{pca} and T_{mda}) and to formulate the dictionary matrix, A, as explained in Section 3.3.3. Each probe cycle is treated independently in our experiments. The distance scores for the various probe cycles are not combined with other cycles belonging to the same subject ID to perform the classification.

Experiments on inter-capture modality platform

For the first set of experiments (Exp - 1a), the key novelty of BGEI-based approach will be evaluated; the use in a cross-capture modality platform. This is performed on the multi-modal segment of the DGD. The frontal depth sequences are used as the probe, while the side-view sequences are used as the gallery.

To increase the database size and the robustness of the results, an extended version of this experiment (Exp - 1b) is carried out which includes the *nw* frontal sequences from the rest of the DGD as impostors in the gallery. To do this, the gallery and probe data is reversed, with the frontal depth sequences forming the gallery set, and the side-view sequences forming the probe. This brings the total number of gallery subjects up to 37.

Experiments on intra-capture modality platform

The BGEI can also be used in domain specific applications. Therefore, the BGEI is also compared to other gait features in their respective imaging domains to see how well this feature performs. First, we compare the BGEI to the GEV using the main

Experiment	Feature	Gallery	Cycles	Probe	Cycles
1a	BGEI	DGD side	8×5	DGD front	8×5
1b	BGEI	DGD front DGD nw	2×5 35×5	DGD side	8×5
2a (*nw-nw*)	GEV / BGEI	DGD *nw*	35×3	DGD *nw*	35×2
2a (*nw-fw*)	GEV / BGEI	DGD *nw*	35×5	DGD *fw*	35×5
2a (*nw-sc*)	GEV / BGEI	DGD *nw*	35×5	DGD *sc*	35×2
2a (*nw-fc*)	GEV / BGEI	DGD *nw*	35×5	DGD *fc*	35×2
2a (*nw-bc*)	GEV / BGEI	DGD *nw*	35×5	DGD *bc*	35×2
2a (*nw-ns*)	GEV / BGEI	DGD *nw*	35×5	DGD *ns*	35×4
2b (*nw-nw*)	GEI / BGEI	CASIA *nw*	124×4	CASIA *nw*	124×2
2b (*nw-cl*)	GEI / BGEI	CASIA *nw*	124×4	CASIA *cl*	124×2
2b (*nw-bg*)	GEI / BGEI	CASIA *nw*	124×4	CASIA *bg*	124×2

Table 7.1: Experiments on cross capture modality platform.

set of the DGD, containing 35 subjects. An intra-class test is performed using the nw sequences. Three of the five *nw* cycles for each subject are assigned to the gallery while the remaining sequences are used as the probe. Inter-class tests are also performed, with all five *nw* cycles forming the gallery, and all available cycles in each of the other classes (*fw*, *sc*, *fc*, *bc*, *ns*) forming the probe in their respective experiments.

A similar set of experiments is performed on the CASIA dataset B, in which the BGEI is compared to the GEI. The dataset contains 124 subjects with three different walking classes: normal walk (*nw*), bag (*bg*) and clothing (*cl*). Six sequences for each subject exist for *nw*, with two for each of the other classes. The experiments on this dataset follow the evaluation outlined in [141]. The intra-class experiment is performed on the *nw* sequences, with four allocated to the gallery and two to the probe. In inter-class tests, again four cycles from *nw* are used as the gallery, while the two sequences in each of the other classes make up the probe in their individual experiments. A summary of all the experiments is detailed in Table 7.1.

7.4.2 Results

To perform the experiments, corresponding features for each of the test cases are computed and transformed to the energised discriminated domain using the learned PCA and MDA basis. Similarity distance is computed using the SRC-based classifier and ROC curves and rank scores are plotted.

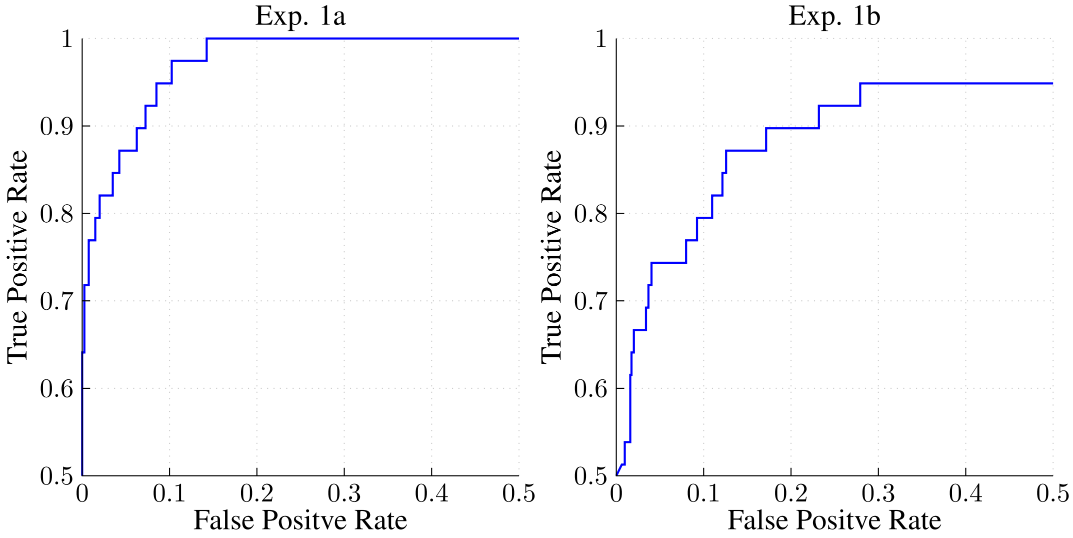


Figure 7.5: ROC curves for BGEI cross-capture modality experiments.

A rank-1 accuracy of 100% is achieved in the first cross-capture modality experiment. The ROC in Figure 7.5, however, does show some miss-verification at low false positive rates. This initial experiment shows great potential for the proposed system, though it is limited to a small dataset of eight subjects. The accuracy is not likely to hold in the expanded experiment, because the larger gallery size would make identification and verification more difficult. The rank-1 accuracy drops to 88.5% in Exp 1b, though 100% accuracy can still be achieved at rank 2. Overall, these results are promising, and demonstrate that performing appearance-based gait recognition using the BGEI across frontal-depth and side-view images is possible.

In the intra-capture modality experiments, an overall drop in the performance of the BGEI has been observed, compared to the GEI and GEV (Figures 7.6 and 7.7). While good results can still be obtained, a fairly notable drop can be seen in the experiments compared to the GEV. This can be attributed to the significant loss of information in the transition from a 3D representation to a 2D one; the separate motions of the left and right legs are no longer retained, and the entirety of the back legs in the gait cycle is lost in order to construct the BGEI. However, the lowering of performance does not exceed 5% at a false alarm rate of 10% indicating that even though the GEV uses this domain-specific 3D appearance information, the BGEI retains significant distinguishing power.

The BGEI fares better against the GEI feature (Figure 7.7), likely due to the fact that less information is lost going from the GEI to the BGEI than from the GEV. Overall, these results on the CASIA dataset are in fact quite similar (the BGEI achieves a

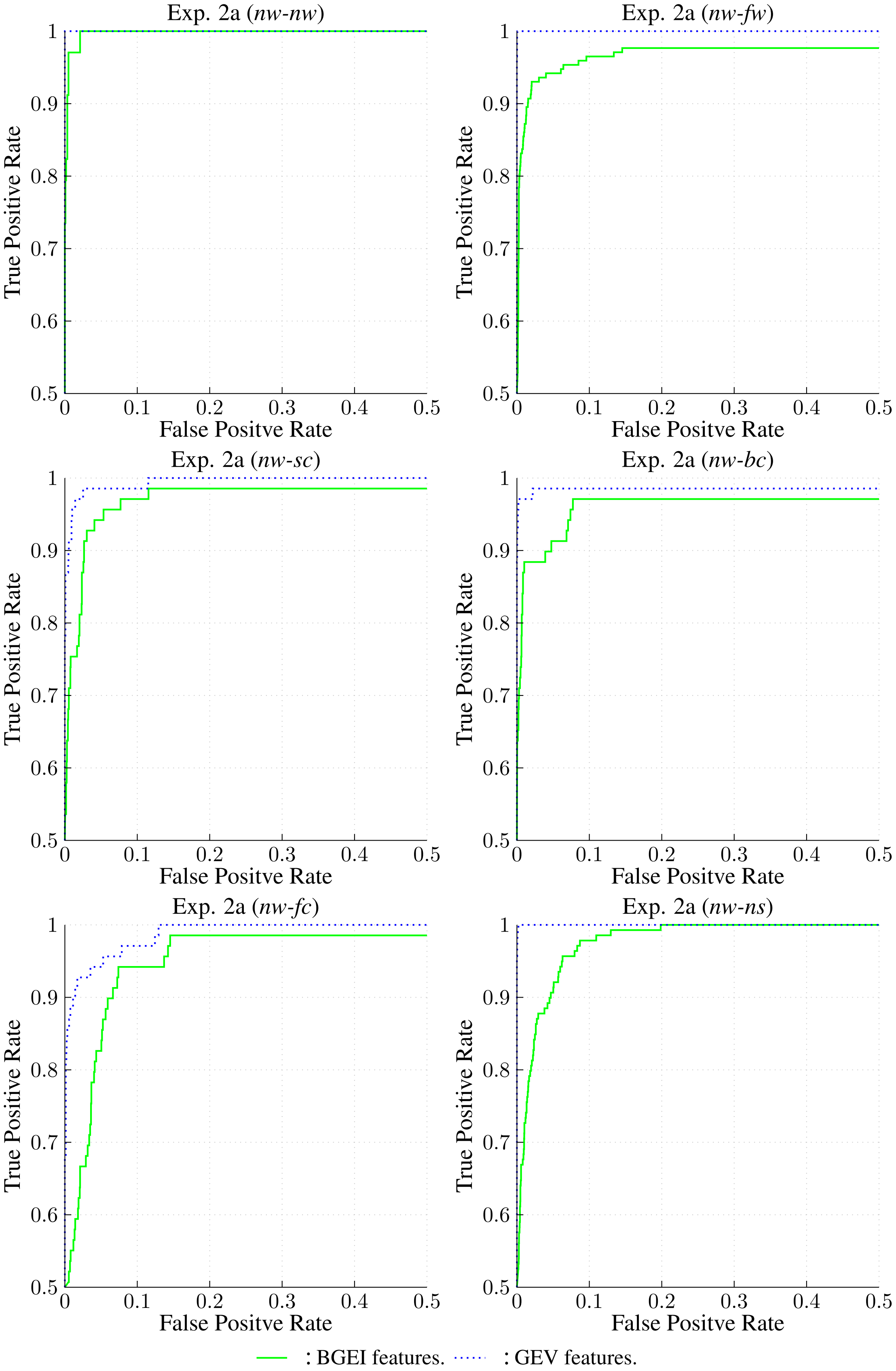


Figure 7.6: ROC curves comparing BGEI features on the DGD database (tests in Exp 2).

	nw-nw		*nw-cl*		*nw-bg*	
	Rank-1	TP @ FAR 3%	Rank-1	TP @ FAR 3%	Rank-1	TP @ FAR 3%
SRC-MDA on GEI	**100.0**	**99.57**	**80.3**	**85.1**	68.5	68.6
SRC-MDA on BGEI	95.0	94.42	79.5	80.8	69.4	73.1
SRC on GEI	97.5	99.14	72.1	79.7	53.2	56.7
LDA on SEIS [165]	99.0	–	64.0	–	**72.0**	-
NN-MDA on GEI [98]	99.0	–	60.0	–	22.0	–
CAS [141]	97.6	–	32.7	–	52.0	–
KPCA on GEI [99]	90.0	–	–	–	–	–

Table 7.2: Recognition performance of BGEI. Recognition performances (CMS at rank-1 and ROC at false alarm rate of 3%) of the proposed algorithms and other reported results in literature on the CASIA database.

slightly lower accuracy in *nw-nw* and *nw-cl* tests, but higher in the *nw-bg* test), showing that the loss of the back contour and back leg does not severely impede its ability to discriminate between different people under those conditions.

Some of these results may also be attributed to the use of the proposed classification method. Table 7.2 lists the rank-1 accuracy and the true positive rate at a FAR of 3% for the experiments we obtained on the CASIA dataset. Listed also, are the results that have been reported by other researchers on this dataset. The proposed classifier significantly improves upon the systems that share the common GEI feature, such as nearest neighbour with MDA [98] or KPCA [99] classifiers. An improvement over modified gait energy features such as the SEIS [165] also have been achieved, though the proposed approach is slightly worse in the *nw-bg* case.

7.5 BGEI for View Invariant Gait Recognition

View invariant solutions proposed in earlier sections of this thesis can only work in a multi-view camera environment with depth cameras capturing frontal depth and RGB cameras capturing the side-view. However, the proposed approaches won't perform well in recognising an individual who is captured by a CCTV camera in an arbitrary view. This is where gait recognition performs very poorly and only 10-30% recognition rate is achieved with significant pose changes [140].

There are two main challenges for view invariant gait recognition systems to perform: (a) the view dependent nature of the extracted gait feature (b): improper inter-

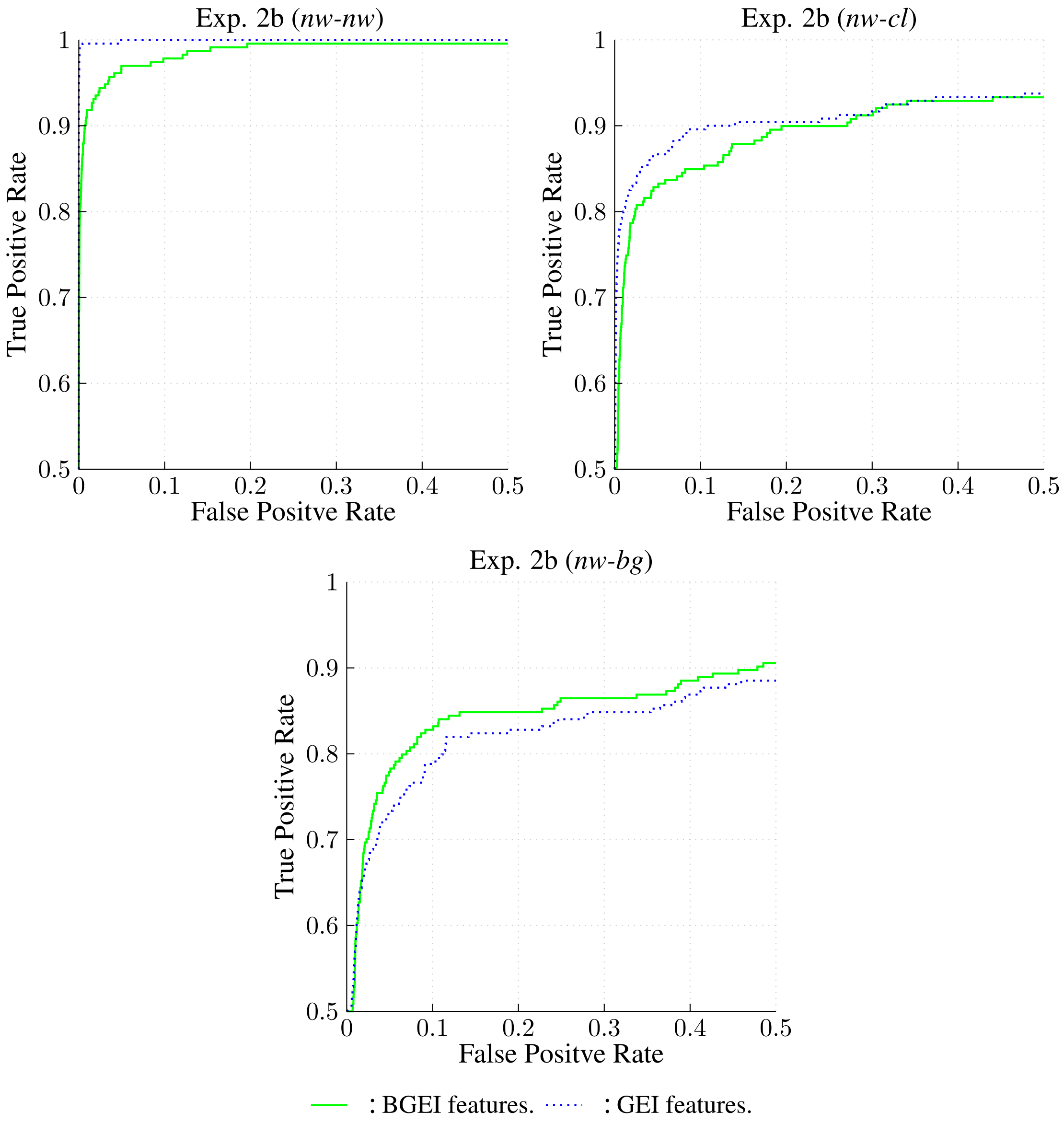


Figure 7.7: ROC curves comparing BGEI features on the CASIA database (tests in Exp 3).

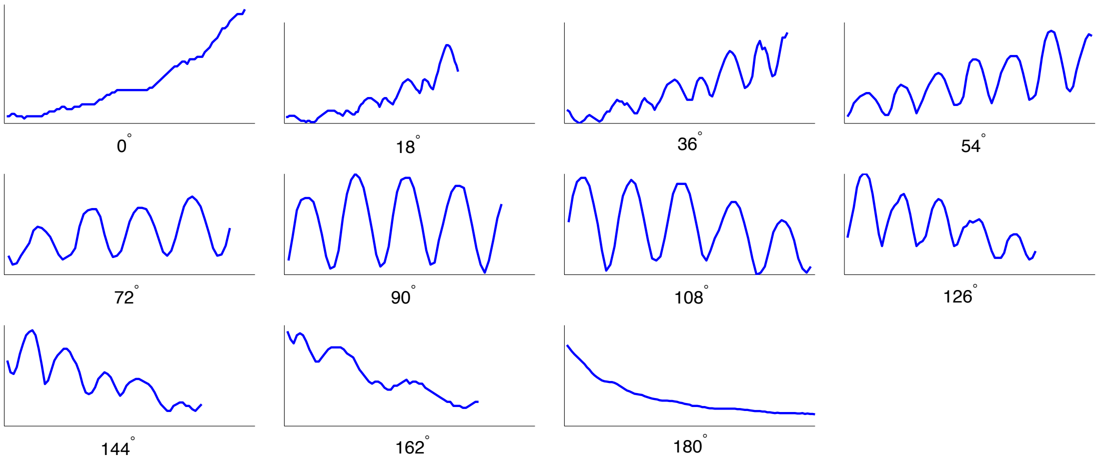


Figure 7.8: Width profiles from multiple views on the CASIA database.

view registration.

There are a number of approaches proposed to resolve these issues particularly using transformational models. View transformation models (VTM) based on the GEI proposed by Worapan *et al.* [140, 166] adopt singular value decomposition (SVD) to transform models to different viewpoints. However, it provides an adequate platform to match different views by transforming them to a common domain; it struggles to find the optimum transformation model due to the significant variations in the projective spatial transformation on the input feature for VTM learning.

In this section, the effectiveness of the proposed BGEI gait feature is explored to provide a view invariant solution.

7.5.1 Gait Cycle Detection for Different Views

The initial step of any gait recognition algorithm is to detect the gait cycle. In the side-view it can be easily done using the smooth oscillation of width profile and it is also not required to detect exact same stances of front and back leg profile as it is approximately symmetrical. However, it becomes harder in other views when gait cycles are not accurately distinguishable from a noisy width signal. Figure 7.8 shows example width profiles computed for various view-angles on the CASIA database.

A robust method has been proposed to accurately segments the gait cycles with left and right leg separation. The proposed method is optimised for each view and enable correct matching of gait cycles between the views. “Double support” phase of gait cycles are quite distinguishable in each frame of silhouettes and it is used for gait cycle

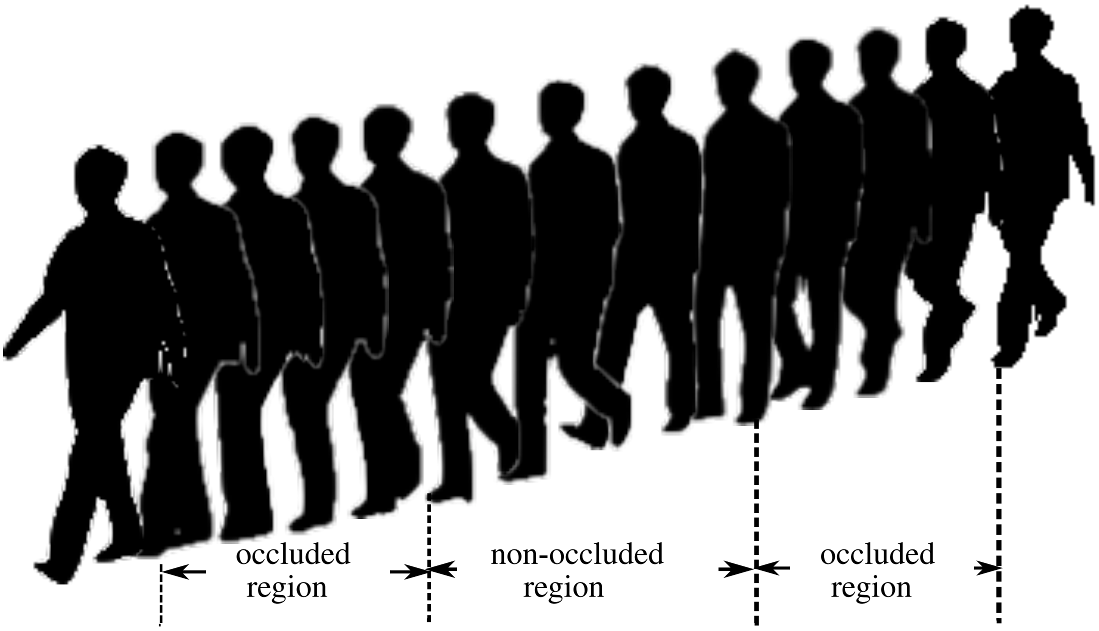


Figure 7.9: Separation of occluded and non-occluded leg region. Examples of 36° view angle silhouettes from the CASIA database.

separation. To detect them, silhouettes on each frame are cleanly extracted and the lower half of the silhouettes is explored for further analysis.

As a first step, front and back legs are separated from the silhouettes. Two different approaches are used to do this, based on occluded and non-occluded positions of legs. Occluded and non-occluded frames are defined by analysing the connected components of 30% of the height from the bottom as shown in Figure 7.9. For occluded legs, front and back separation are done based on the horizontal centre of the lower body and for non-occluded legs, connected components are detected and front and back legs are assigned based on their relative position to the horizontal centre.

For the non-occluded regions, the toe position of each separated leg is computed by finding the extreme lower points at the bottom. To make it robust towards different ankle positions, appropriate rotation matrix is applied to the leg portions based on the pixel density at bottom as shown in Figure 7.10. Since the main concern is to find the “double support” phase, which mostly occurs in non-occluded regions, the simple mean value of extreme lower leg points is used to determine toe positions in occluded regions.

To determine double support-positions and left and right leg separation from the ground point positions, an alternative of occluded and non-occluded is used. To avoid

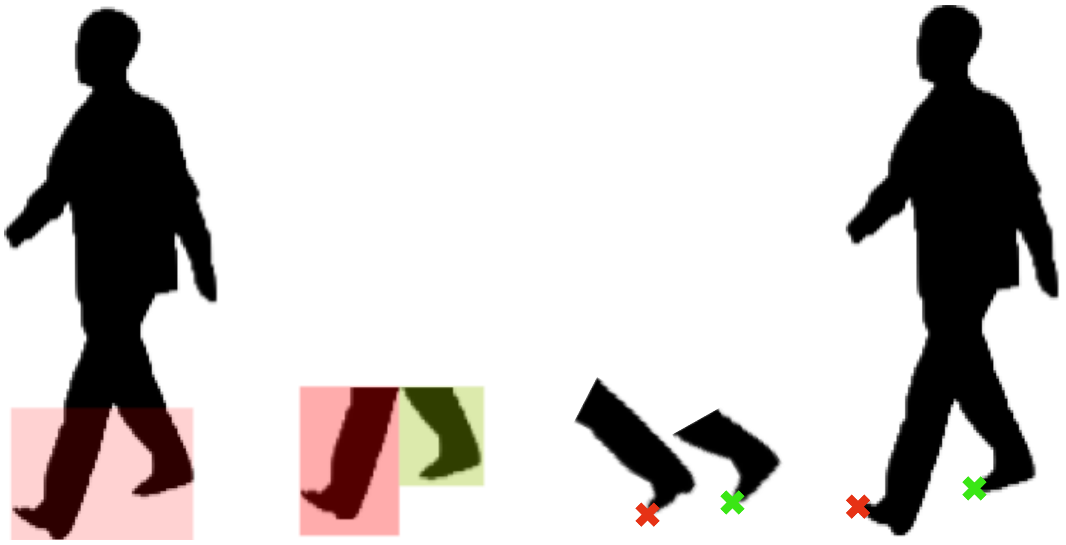


Figure 7.10: Toe detection for non-occluded legs (Bottom Leg separation, rotation of separated bottom part based on pixel density and toe detection based on the extreme points).

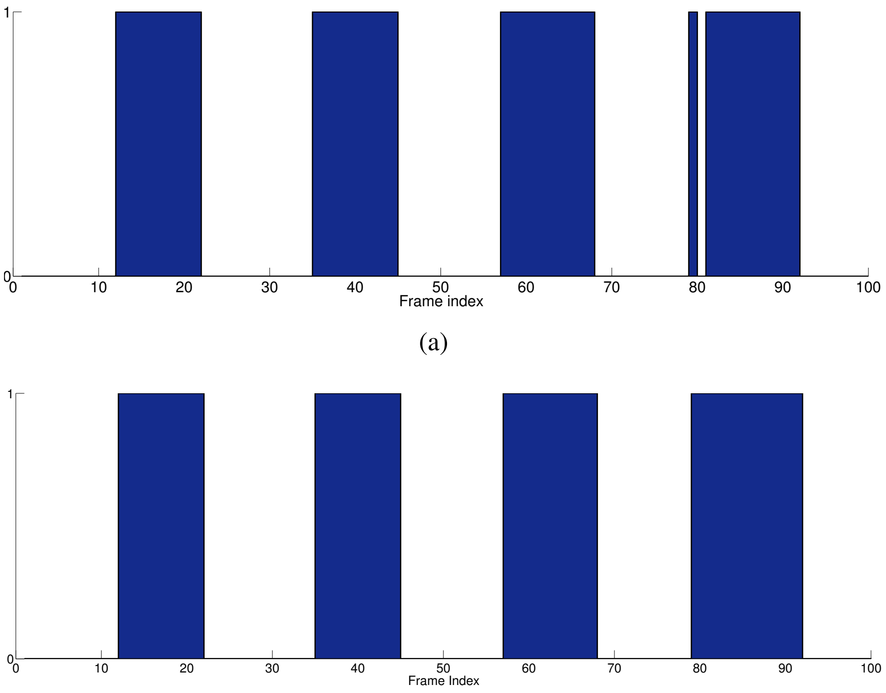
the outlier due to the segmentation noise that causes the noisy alternatives, a moving window of size five frames is used to smooth the signal as shown in Figure 7.11.

After correct alteration has been found with the accurate toe detection in non-occluded regions, alternative double supported striking points with maximum stretch are chosen to segment the gait cycles as shown in Figure 7.12.

7.5.2 View Invariant BGEI

At the beginning of this chapter, it has been shown that back leg information is just replication of the front leg gait pattern due to the symmetrical nature of human anatomy, and acceptable recognition performances can be achieved using the BGEI feature. This allows the BGEI to be applied to gait features in the presence of view angle changes as the frontal contour of the walking silhouette is similar nature with wide view angle changes.

To explore the applicability of BGEI to a view invariant gait recognition system, silhouettes from multi-view images are extracted and gait cycles are computed, as explained in Section 7.5.1. These silhouettes are further pre-processed to enhance the inter-view registration by projecting them to a common view angle (90° view is chosen as the common view angle for better gait representation). Then, these images



(a)

(b)

Figure 7.11: Separation of frames where occluded and non-occluded leg regions occur. In (a) and (b), frames that have been detected as containing occluded leg regions, get one and the others zero. In (b) moving window with frame size of five is used to smooth the outliers.

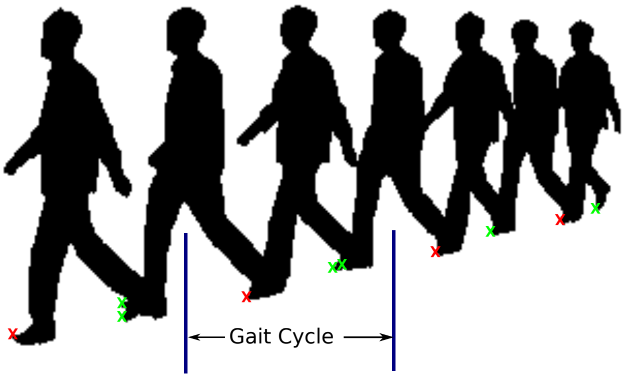


Figure 7.12: Extreme double support points detection and gait cycle segmentation. Examples from the view angle of 36° on the CASIA database are shown.

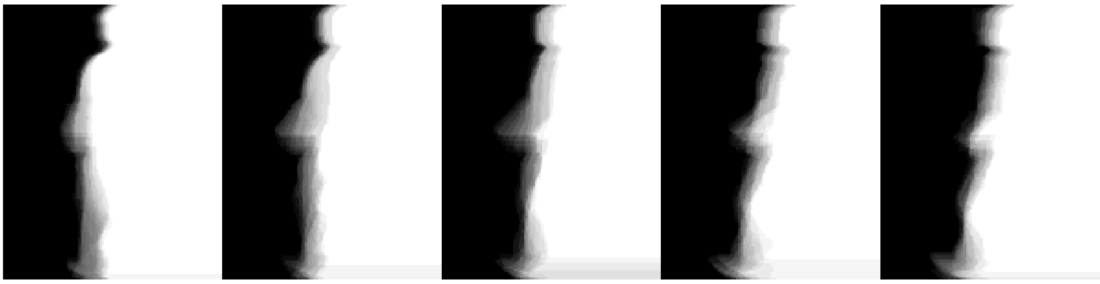


Figure 7.13: Example BGEIs from multi-view images on the CASIA database. From left to right BGEIS generated from view angles of $18°$, $36°$,$54°$, $72°$ and $90°$ are shown.

are scaled to the common height and horizontally aligned to the horizontal centre of the frontal contour. Then, these aligned silhouettes are backfilled from their frontal contour and averaged to compute the BGEI explained in Section7.3. Example BGEI images computed on multi-view images on the CASIA database are shown in Figure 7.13.

From Figure 7.13, it can be observed that the BGEI still struggles to perform proper registration when the view angle changes by more than $36°$. This explains that BGEI needs to be either enhanced for proper registration or needs to be transformed to the common basis domain. To do this, a singular value decomposition based view transformation model similar to the method in [140] is chosen to learn the transformation basis from a particular view to the view on the enrolled data as it produced promising recognition performance in [140] on GEI features.

At first, the BGEI features from the training data are wrapped into a column vector to form a dictionary, D. For the robust results of SVD, the input feature need to the most energetic feature that corresponds to the particular individual in particular view. This is achieved by learning a local PCA basis (T) from the training data for each view angle. The locally learned PCA basis is chosen, as the generalised PCA is biased towards the perspective information of different view angles. Training features on the dictionary, D, are transformed to the local PCA domain using the learned bases as follows,

$$g_v = T_v * D_v, \tag{7.1}$$

where v corresponds to view angle, g_v is the transformed feature vector of the training data D_v. SVD is learned on this transformed feature as shown in Equation 7.2,

$$\begin{bmatrix} g_1^1 & \cdots & g_1^I \\ \vdots & \ddots & \vdots \\ g_V^1 & \cdots & g_V^I \end{bmatrix} = USV^T = \begin{bmatrix} T_1 \\ \vdots \\ T_V \end{bmatrix} = \begin{bmatrix} x^1 & \cdots & x^I \end{bmatrix}. \tag{7.2}$$

The dictionary shown in the left of Equation 7.2 contains transformed feature vectors of individuals and views. Each column represents the feature for a single individual and each row belongs to a particular view. U is the orthogonal matrix with the dimension of $Vn \times I$, where n is the feature dimension. S is the $I \times I$ diagonal matrix containing singular values. T_v is the projection matrix that can project intrinsic gait feature vector x^i to the specific view angle v and it is independent of the subject. Using the learned projection matrices, the BGEI features can be transformed from angle a to b as follows,

$$g_a = T_a T_b^+ g_b, \tag{7.3}$$

where T_b^+ is pseudo inverse matrix of T_b.

After transforming the features into the single view domain, multiple discriminant analysis is used to learn the discriminant basis that increases the inter-subject variance, while reducing the intra-subject difference. Here, MDA is applied as a post-processing step in contrast to the one used in [140] to reduce the effect of outliers that can be caused by view differences.

7.5.3 Experiments and Results

The effectiveness of the view independent nature of the BGEI feature is evaluated on the CASIA database on the view angles of $18 - 162°$. For training, the first four GEIs of normal walk are used and the remaining two are used as the probe. Figure 7.14 compares the Rank 1 cumulative scores using BGEI and GEI features. It can be noted that around $15 - 25\%$ improvement in identification rate has been obtained for larger view angle changes using BGEI features compare to GEI features as used in [140]. However, recognition rate is comparatively lower in both cases when the view angle deviation exceeds 36°. It has also been noticed that in intra-view angle cases, GEI shows higher performance compared to BGEI. This is attributed to loss of gait information on BGEI due to the back leg. However, this loss performance degradation is negligible compared to the improvement achieved in inter-class cases. The above analyses conclude that BGEI is a sophisticated feature to perform gait recognition when the subject's view angle deviates with the following limitation: BGEI still needs a view transformation model for adequate inter-view feature registration, lower performance in intra-view and recognition performance achieved in more than 36° is still below 50%.

Based on the initial investigation, the following future directions are proposed to solve the above issues,

- Synthesis 2D view from pre-enrolled 3D model
  Registering individuals to a 3D model and synthesising a 2D image from them to match with an arbitrary view is a promising direction to perform view transformation without any model training. In the enrolment process, it is feasible to capture an individual in 3D by using a multiple camera setup or depth sensors. As camera position of the surveillance camera is known, the view angle of a walking individual can be computed by analysing the walking direction [167]. Using the walking direction and camera positions, 3D to 2D projective transformation can be applied to generate the synthesised 2D view of a particular person from the enrolled 3D volume. Then, features extracted from these two images can be used to compute the match score for classification.

- Image registration before view-transformation
  Apply image processing techniques such as projective transformation using the known land marks before applying transformation models. Since transformation models are sensitive to the spatial position of the silhouette, it is important to ensure the appropriate inter-view registration as much as possible before applying any transformational models on the feature.

- Explore different types of view transformational model
  There are several transformation models that are used in the signal processing domain including MDA, PCA, SVD, DCT, sparse representation-based learning model, neural network learning model etc., each having their own strengths and limitations. It also necessary to explore them on the stage on which these transformation models have to be applied (global space or local space) and in which combination.

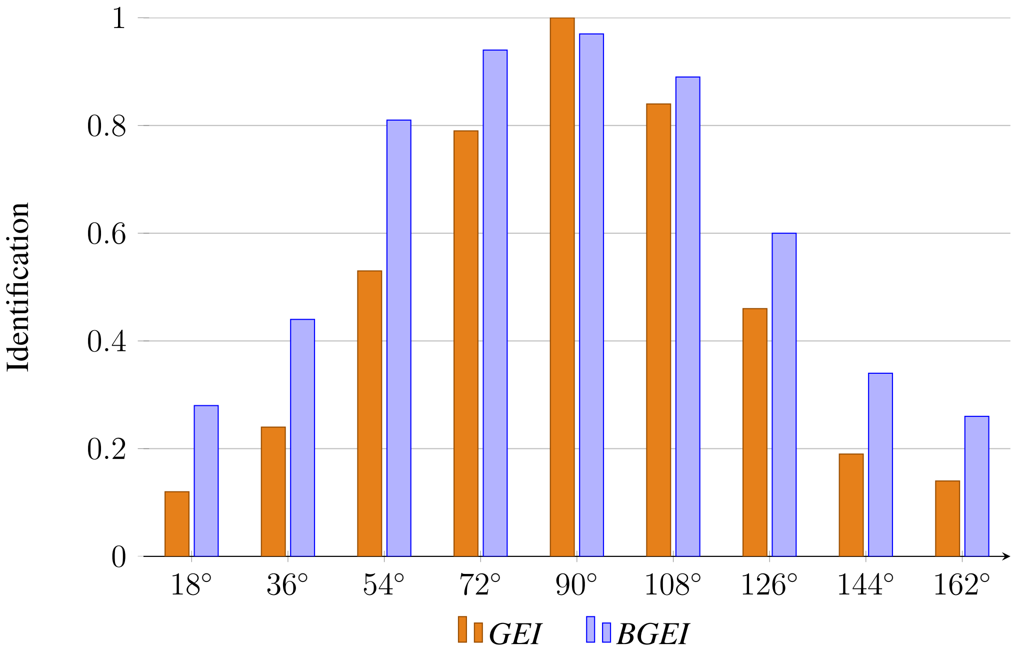


Figure 7.14: Rank-1 cumulative scores comparing BGEI and GEI features for different views.

7.6 Summary

In this Chapter, a novel approach has been proposed for gait recognition that enables the use of multiple independent capture sources within a single gait recognition system through a feature, the BGEI, that can be synthesised from multiple input sources such as frontal-depth or side-view data. Experiments show that the proposed BGEI has the potential to work in a cross-capture platform with the initial results of a CMS of 100% at rank-1, albeit on a small database. Performance of the BGEI can be further improved by incorporating advanced spatio-temporal alignment and scaling between the cross-capture platform.

The construction of the BGEI discards information, most notably, the entirety of the back leg, that is retained as domain-specific features such as the GEI and GEV. However, through an evaluation on the CASIA database, it has been demonstrated that there is sufficient information in the frontal plane of motion to recognise subjects at comparable and sometimes even greater accuracy, compared to the traditional GEI.

Finally, the benefits of applying BGEI to the view invariant system are explored using a PCA-SVD-MDA transformation matrix, and future directions from synthesising 2D from 3D approach are proposed.

Chapter 8

Conclusions

Gait is a relatively new and emergent biometric that pertains to the use of an individual's walking style to identify or validate an identity of an individual [168, 169]. The potential of gait as a biometric has further been encouraged by the considerable amount of evidence available across several fields, especially medicine [48]. The unique advantage of gait as a biometric is that it can be used for passive identification of people, (i.e, it can be used at a distance or at low resolution). This provides strong potential for gait recognition systems to yield benefits to forensic analyses in surveillance environments and to portal-based smartgate security applications. However, gait recognition systems have poor intra-person reliability due to their dynamic nature and are dependent on various physiological, psychological and external factors such as footwear, clothing, surface of walking, mood and illness. This thesis has focused on improving gait recognition performance in more unconstrained challenging conditions and in particular, it has developed new solutions to incorporate gait recognition into real-world applications. Throughout the various analyses using signal processing, machine leaning, statistical learning and computer vision techniques, the following contributions are made:

- Silhouette segmentation on gait recognition performance is explored and the influence of silhouette spatial regions and dynamic-static nature on appearance-based features for distinguishing individuals has been experimented.

- Improved gait recognition performance with the solution to view dependency and self-occlusion by the proposed fast and efficient quasi model-based ellipsoid fitting and appearance-based 3D feature, GEV.

- Introduced patch-based feature optimiser that can handle the unwanted variations in the image, whether they be due to lighting or pose.
- Proposed optimised SRC-based classifier on gait recognition context that achieves state-of-the-art recognition performance.
- Improved frontal gait recognition that enables incorporation of gait recognition into portal-based security applications.
- Gait recognition that can perform in cross-capture modality systems.
- Initiating steps to the view invariant gait recognition in 2D.

The research has been initiated by improving the recognition performance of the well-performed and most popular baseline algorithm, by exploring several influencing factors on appearance-based gait recognition. By comparing the baseline performance on silhouettes resulted from background subtraction and the cleaned silhouettes resulted from enhanced graphcut-based motion segmentation, it has been concluded that silhouette appearance is purely contributing to gait recognition rather than the artificial static noises generated from clothing and carrying goods. In addition to that, 5-10% improved rank 1 recognition rates are obtained on inter-class test cases.

Further investigation has been completed to improve the inter-class recognition performance. The effect of clothing and carrying goods on dynamic and static features and spatial position has been explored. Based on the outcome, it has been concluded that dynamic regions are less susceptible to the inter-class appearance changes and only belong to the lower region of the silhouette. With this conclusion and the spatial analysis on GEI, improvement of 20-30% of recognition rate has been achieved using only 40% of the GEI.

Since the algorithms proposed in 2D are view dependent and perform best when a side view is used, the research work has been extended on 3D gait recognition by reconstructing 3D voxel volumes on the CMU MoBo gait database. Using these reconstructed voxel models, derived from silhouettes from multiple views, a novel, semi-model-based gait recognition algorithm is proposed to extract Ellipsoid parameterisation of the voxel model. The proposed algorithm addresses the limitations in the model-based and appearance-based techniques by combining the strengths of each approach. The fitting of ellipsoids to a voxel model performs much faster than full pose estimation systems, yet still provides a direct estimate to the underlying kinematic features (joint angles). The move to 3D space solves the issue of view dependency and problems with self-occlusions, though our method does constrain its applications to situations where a multi-camera setup is in existence. While we demonstrate the proposed ellipsoid fitting approach to perform gait recognition, it should be noted that the model could also be used in other applications such as gesture recognition. The use of 3D information also allows left and right legs to be easily segmented, and gate cycles to be detected. The proposed approach achieves 10-15% improvement in recognition performance over its 2D counterpart, particularly when there is a class mismatch between the gallery and probe sequences. These outcomes were published in [170].

Though the 3D-model-based approaches solve the view dependency and also self-occlusion, the recognition performance achieved is comparatively lower than the side-view appearance-based techniques such as GEI, due to the inability of proper model fitting on the available data. Therefore, an extension of the GEI is proposed to operate in the 3D domain, using binary voxel volumes instead of 2D silhouettes. This proposed GEV algorithm shows an improvement over its 2D variant in performing gait recognition, given multi-view data.

Having a multi-view camera setup, however, can be impractical under many applications, even in controlled conditions such as gait-based biometric authentication. An alternative to acquiring this 3D data would be to use some form of depth sensing device. The applicability of GEV has also been demonstrated on frontally captured depth images, such as would be acquired using a biometric portal, through the use of synthesised data from reconstructed voxel-volumes on the CMU MoBo database. Though, frontal GEV loses the back leg information, the synthesised frontal GEV features out-perform all the baseline, including frontal, side, multi-view GEIs and even the full GEV. This is likely due to essentially all gait characteristics being acquired from a frontal perspective, while the back filling reduces the impact of noise. Frontal based depth has the advantage of being able to capture essentially all characteristics of gait from a single viewpoint without the issue of self-occlusion. These depth images can be easily acquired using devices such as Microsoft Kinect. A frontal viewpoint also makes it possible to easily integrate into biometric portals.

Inspired by the results obtained in synthesised frontal volumes, the applicability of frontal GEV have been further explored on real partial volume reconstructions from depth images captured from a frontal viewpoint. To facilitate this, the depth gait database (DGD) is proposed. Evaluation of the frontal GEV on DGD database with 37 subjects achieved more than 95% Rank-1 recognition performance in all the test cases. This pioneering work on 3D appearance feature and frontal GEV using the Microsoft

Kinect has been published in [171].

The performance of the baseline, GEI, is more sensitive to the changes in appearance. To minimise the effect of appearance changes on GEI, patch-based histogram descriptors have been explored and a novel descriptor, the histogram of weighted local directions (HWLD), has been proposed to tolerate the appearance variations. The proposed HWLD features demonstrate state-of-the-art performances, showing average of 10% improvements in inter-class test cases in the CASIA dataset. Superior performance of the feature on the high population OULP dataset, which contains more than 3000 subjects, shows that the proposed method is stable over a large population. Furthermore, we demonstrate that local histogram feature extraction techniques are much more stable to minor segmentation errors. They also show improved performance in the absence of feature conditioning processes such as PCA and/or MDA in inter-class tests.

The advantage of the proposed feature optimisation algorithm, HWLD, is that it can be extended to 3D without any exponential increase in feature size. Extended 3D version of the HWLD is proposed, with evaluations using GEVs on the DGD dataset beating all known results. This HWLD-based solution for the appearance-based gait features was published in [172].

The SRC based classifier is explored in a gait recognition context and the benefits of applying SRC to a MDA-based discriminated input space has been shown through a comparison with the traditional PCA-kNN and SRC-PCA based classifiers. The significant improvement of the proposed method over the recent approaches shows that discriminating the input space, based on the local class-labels and applying an SRC based classifier is the optimum solution for future classification tasks. These outcomes were published in [172, 173].

A novel approach for gait recognition that enables the use of multiple independent capture sources within a single gait recognition system is proposed. The introduced feature, the backfilled gait energy image (BGEI), can be generated from frontal-depth and side-view data. To allow the cross-capture modality evaluation, DGD is extended with sequences captured from the side as well. It has been shown that the proposed BGEI has the potential to work in a cross-capture platform with the initial results of a CMS of 100% at rank-1, albeit on a small database. The construction of the BGEI discards information, most notably, the entirety of the back leg, that is retained in domain specific features, such as the GEI and GEV. However, through an evaluation on the CASIA database, we have demonstrated that there is sufficient information in the frontal plane of motion to recognise subjects at comparable, and even sometimes greater, accuracy, compared to the traditional GEI. This cross-capture modality solution is published in [56].

8.1 Future Research

This thesis has contributed to several areas of a gait recognition system, however there are still many areas that future work could address. Areas of future work, that further improve techniques proposed in this thesis, as well as potential new research are listed below:

- **Approaches for view invariant gait recognition on arbitrary view angle in 2D:**
  The backfilled gait energy image provides a direction to explore a view-independent feature to perform gait recognition in an arbitrary view-angle in 2D. However this BGEI needs to be further enhanced for optimised registration for large view angle deviation. Camera calibration details and walking direction can be incorporated in the analysis of inter-view registration to perform perspective transformation. Different types of view transformational models are also can be explored in global and local space, for combined performance.

  A synthesised 2D test image from a registered 3D image also can be explored to verify the test subject in arbitrary view angle. Since there are depth sensors and 3D multi-view set ups that are possible during the enrolment phase, it will be a promising direction to register the subjects in 3D. During the verification process, using the camera orientation and walking direction, a 2D feature can be synthesised from the 3D enrolled volume. This synthesised feature can be used to match with the test subject's feature.

- **Experiments on large scale data with common evaluation protocol:**
  To ensure the gait recognition system's applicability in the real world environment it needs to be evaluated on large scale data. At this time, the maximum database available to evaluate gait recognition algorithms is around 3800-4000. However, it needs to be increased to large scale with appropriate evaluation protocol for the reliable estimation of recognition performance and to make the common scale for comparison.

- **Fusion with the other biometrics:**
  Face and iris are the commonly used biometrics in portal-based smart-gate applications. A gait recognition system can be developed to integrate with these biometrics to provide a non-intrusive verification process and higher accuracy. Best performing fusion strategies can be explored for identifying an optimised fusion method (sum, average, weighted sum *etc.*) and to recognise on which level fusion needs to be performed (feature level or score level).

- **Enhanced sparse representation-based classifier:**
  The objective function of the sparse representation-based classifier used in this thesis is based on the clean nature of the feature arrangement, that doesn't include any penalty functions to include outliers and noisy features. Recently there are number optimised objective functions that have been introduced and have performed well in object classification [174]. It is worthwhile evaluating SRC using the penalty-based objective function with fine-tuned optional parameters in gait context.

- **Deep networks for classification:**
  Deep networks can compactly represent a significantly larger set of data in compact form. A deep network can have significantly greater representational power and compute much more complex features of the input. Since it represents non-linear functionality, it can better separate an individual characteristic from the outlier. Because of these advantages, it's a highly recommended direction to explore deep networks for better classification of gait features.

Appendix A

Experimental Results

In this Section, comprehensive details of the recognition performance matrices evaluated on several experiments are listed for reference.

Comparison of SSD and ISD on the GEI-baseline

Table A.1 compares the variation in recognition performance using the two different scoring methods, single score similarity distance and independent score similarity distance. Details of these approaches are explained in Chapter 3.

Score	Exp	ROC				CMS		
		TP@ 1%	TP@ 3%	TP@ 5%	AUC	Rank1	Rak3	Rank5
SSD	*nw*	1.000	1.000	1.000	0.999	100.00	100.00	100.00
	nw-cl	0.218	0.298	0.323	0.753	45.161	63.710	67.742
	nw-bg	0.145	0.226	0.282	0.687	32.258	53.226	64.516
ISD	*nw*	0.988	0.996	0.996	0.999	99.194	99.194	99.597
	nw-cl	0.19	0.270	0.315	0.747	46.371	61.290	65.323
	nw-bg	0.149	0.230	0.278	0.686	33.468	52.823	63.306

Table A.1: Comparison of single score for similarity distance (SSD) and independent score for similarity distance (ISD) on GEI-baseline recognition performances on the CASIA database.

Influence of Segmentation on Gait Recognition performance

Figure A.1 compares the influence of segmentation on the CMU MoBo database using GEV features. Tables A.1(a) and A.1(b) shows the area under curve (AUC) scores using GEI baseline on the CASIA database and GEV on the CMU MoBo database.

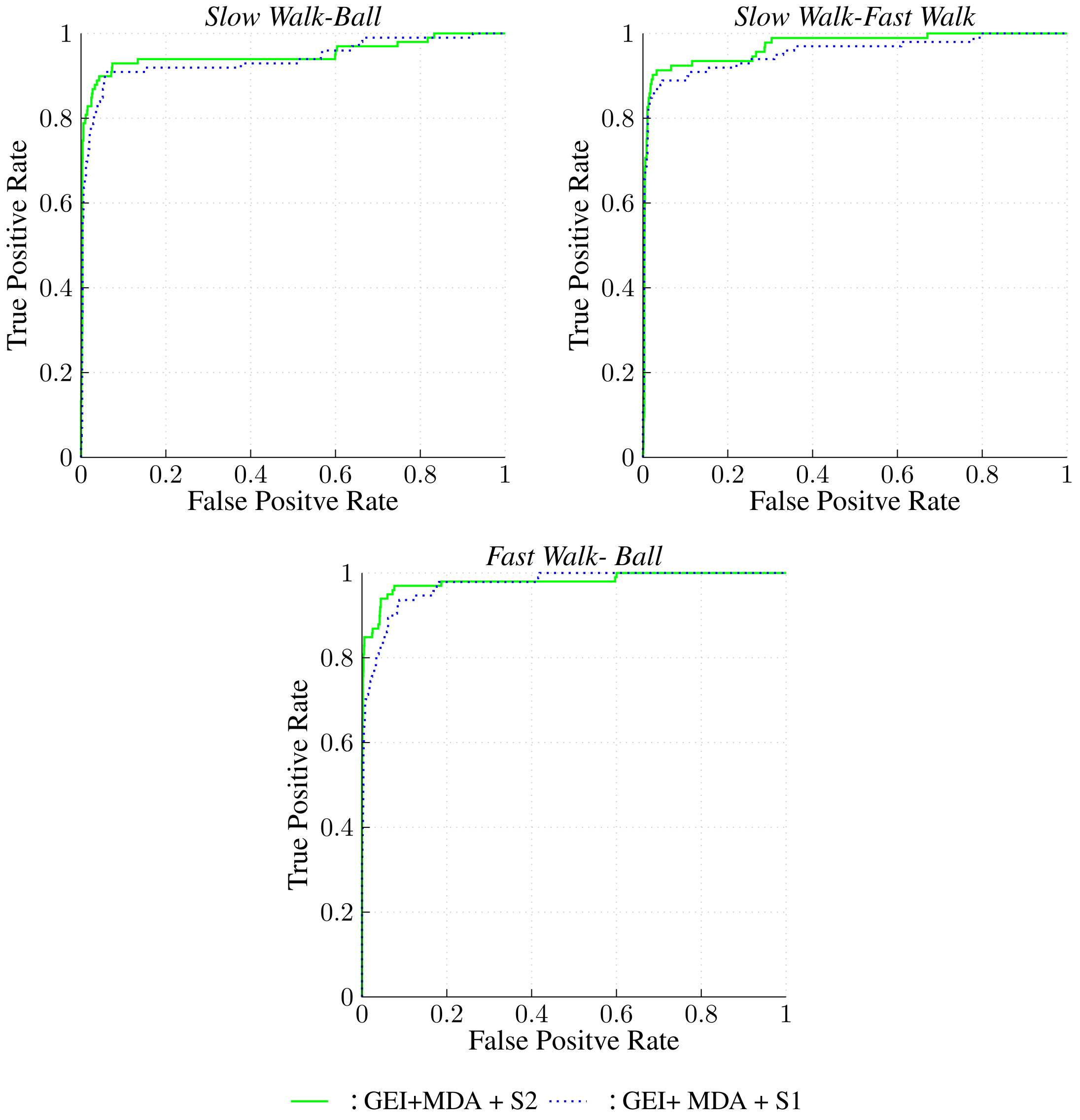


Figure A.1: ROC curves comparing GEI+MDA using different segmented silhouettes on the CMU MoBo database.

(a) CASIA AUC.

	nw-nw	*nw-cl*	*nw-bg*
Original	0.9957	0.9196	0.8428
Cleaned	0.9991	0.9582	0.9036

(b) MoBo AUC.

	sw-b	*fw-b*	*sw-fw*
Original	0.9726	0.9547	0.9396
Cleaned	0.9800	0.9708	0.9501

Table A.2: AUC results comparing the influence of segmentation. Evaluations on the CASIA database and the CMU MoBo database are shown.

Feature	Seg.	PCAMDA	TAC @ FAR 3%			TAC @ FAR 5%			AUC		
			nwnw	*nwbg*	*nwcl*	*nwnw*	*nwbg*	*nwcl*	*nwnw*	*nwbg*	*nwcl*
HWLD	S2	-	**99.6**	**87.7**	**96.8**	99.6	**91.6**	**97.3**	**0.999**	**0.984**	**0.991**
	S2	✔	99.5	86.7	96.3	99.5	90.3	96.3	0.999	0.980	0.993
	S1	-	98.2	87.8	94.8	98.7	92.5	96.0	0.997	0.984	0.986
LDP	S2	-	98.2	84.9	95.0	99.1	89.1	96.3	0.995	0.961	0.984
	S2	✔	99.0	81.9	92.6	99.0	85.0	94.0	0.998	0.952	0.978
	S1	-	97.8	82.8	94.2	98.6	87.2	98.1	0.992	0.954	0.991
HOG	S2	-	98.6	80.2	95.4	99.1	84.6	95.9	0.999	0.962	0.989
	S2	✔	98.6	78.8	94.0	99.5	84.5	94.9	0.999	0.951	0.987
	S1	-	98.2	80.0	94.2	98.7	85.1	96.2	0.998	0.970	0.993
GEI	S2	-	97.2	58.1	87.6	98.2	62.5	89.3	0.994	0.860	0.954
	S2	✔	99.5	66.5	91.0	**100**	71.5	92.7	0.999	0.920	0.966
	S1	✔	99.1	60.2	82.6	99.6	71.3	88.9	0.999	0.864	0.946
GEI [56]	-	✔	-	-	-	-	-	-	-	-	-
SGEI [117]	-	✔	-	-	-	-	-	-	-	-	-

Table A.3: Verification rates comparing the descriptors. Seg refers to the silhouette set used (S1 or S2). PCAMDA refers to the use of feature modelling (PCA and MDA). Our experiments are compared with other state-of-the-art results found in the literature for this dataset.

Comparison Feature Descriptors

Table A.3 and A.4 compare the verification and identification performances of the histogram-based feature descriptors in terms of segmentation noises and with the PCA-MDA model training on the CASIA database.

Feature	Seg.	PCAMDA	Rank-1			Rank-3			Rank-5		
			nwnw	*nwbg*	*nwcl*	*nwnw*	*nwbg*	*nwcl*	*nwnw*	*nwbg*	*nwcl*
HWLD	S2	-	**100**	**92.2**	**96.5**	**100**	**96.6**	**97.6**	**99.6**	**87.7**	**96.8**
	S2	✔	100	91.3	96.5	100	94.8	98.2	99.5	86.7	96.3
	S1	-	99.1	90.4	94.8	99.1	93.0	97.4	98.2	87.8	94.8
LDP	S2	-	98.2	85.1	93.8	99.2	92.1	97.3	98.2	84.9	95.0
	S2	✔	100	79.2	92.1	100	87.9	94.7	99.0	81.9	92.6
	S1	-	97.4	83.4	93.6	99.2	90.2	98.8	97.8	82.8	94.2
HOG	S2	-	99.1	80.8	96.3	100	91.3	97.3	98.6	80.2	95.4
	S2	✔	99.2	78.3	93.8	100	88.	97.4	98.6	78.8	94.0
	S1	-	98.2	80.2	96.2	99.1	92.2	98.1	98.2	80.0	94.2
GEI	S2	-	98.2	57.4	80.9	100	65.2	91.3	97.2	58.1	87.6
	S2	✔	100	74.1	91.2	100	75.6	92.0	99.5	66.5	91.0
	S1	✔	100	64.3	82.6	100	69.6	87.8	99.1	60.2	82.6
GEI [56]	-	✔	100	68.5	80.3	-	-	-	-	-	-
SGEI [117]	-	✔	99.0	72.0	64.0	-	-	-	-	-	-

Table A.4: Cumulative match scores comparing the descriptors. Seg refers to the silhouette set used (S1 or S2). PCAMDA refers to the use of feature modelling (PCA and MDA). Our experiments are compared with other state-of-the-art results found in the literature for the CASIA dataset.

Appendix B

Camera Calibration

Gait recognition on 3D data solves many issues in 2D gait recognition with the improved richer gait information. In this thesis, state-of-the-art recognition performances are shown using 3D Gait data. However, only little research has been done and reflected in literature due to the limitation of multi-view data. Though the CMU MoBo database and the CASIA databases have been available for long time, there were no attempts made in 3D, since these databases don't have their calibration details. As a part of this research, camera calibration for the CMU MoBo database and the CASIA database is computed to reconstruct the 3D volume as illustrated in Chapter 4. To facilitate the future direction in 3D, the computed calibration parameters are provided in this section. Using these camera calibration parameters, points in the image plane (u, v) and a 3D real-world coordinate system are related as follows,

$$s\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_x & r_{xy} & r_{xz} & t_x \\ r_{yx} & r_y & r_{yz} & t_y \\ r_{zx} & r_{zy} & r_z & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{B.1}$$

CMU MoBo Database

CMU MoBo database is a synchronised multi-view database having image footage from six different cameras positioned as shown in Figure B.1. The intrinsic and extrinsic camera calibration details are shown in Table B.1.

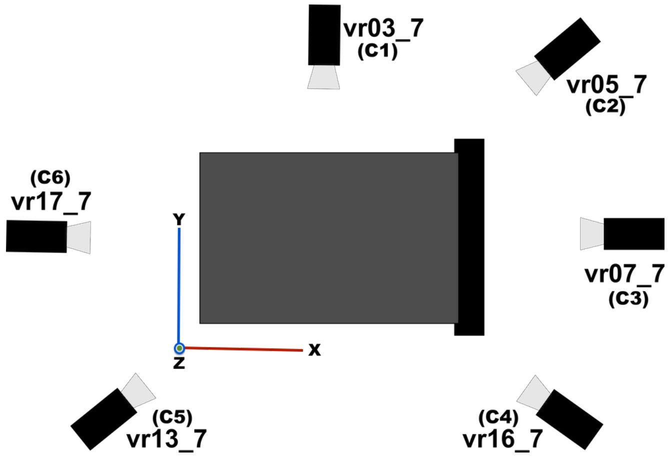


Figure B.1: Camera positions of the CMU MoBo database.

Camera	Yaw	Pitch	Roll	T_x	T_y	T_z	F_x	F_y	C_x	C_y
C1	1.57	3.57	0.07	-7.03	02.44	7.98	839.56	850.52	241.61	318.38
C2	5.53	3.32	0.05	10.37	-5.22	5.30	991.75	948.51	242.34	320.45
C3	0.64	3.33	0.02	-3.88	-6.05	5.23	840.95	843.97	244.64	326.76
C4	-1.54	3.56	0.01	12.56	02.32	7.72	768.54	762.56	233.27	319.01
C5	-2.21	3.34	-0.10	12.29	9.25	5.92	900.23	903.74	239.30	318.27
C6	-3.01	3.40	0.01	3.84	11.30	5.83	804.98	811.4520	247.49	322.00

Table B.1: Camera Calibration of CMO MoBo database [31]. Yaw is rotation respect to Y direction, Pitch is rotation respect to X direction and Roll is rotation respect to Z direction. T_x, T_y, T_z are translation distances, (F_x, F_y) is focal length and (C_x, C_y) is camera centre.

CASIA Database

CASIA database is large scale database with walking individuals captured in 11 views. There are random frame skipping and frame offsets, making this databases hard to use in 3D direction. Cameras are positioned to capture the walking individual from $0°$ to $180°$. That covers the field-of-view of an individual's left hand side. Though the full 3D reconstitution cannot be performed due to the lack of field-of-view, research can be directed based on partially reconstructed 3D data . To facilitate the future research in this direction, camera calibration details are provided in Table B.2.

View	Yaw	Pitch	Roll	T_x	T_y	T_z	F_x	F_y	C_x	C_y
0°	-0.02	3.3	0.01	1.31	-5.5231	1.96	407.34	405.54	160.00	120.00
18°	-0.33	3.33	0.030	3.32	-5.07	2.01	421.92	413.57	160.00	120.00
36°	-0.66	3.26	-0.02	5.29	-3.81	1.87	403.52	419.88	160.00	120.00
54°	-0.96	3.29	-0.03	6.48	-2.23	1.83	379.63	390.95	160.00	120.00
72°	-1.26	3.30	0.02	7.79	-0.59	2.36	439.99	428.22	160.00	120.00
90°	-1.58	3.28	-0.02	7.79	1.55	2.13	422.81	414.71	160.00	120.00
108°	-1.88	3.2745	-0.01	7.13	3.46	1.92	389.32	394.52	160.00	120.00
126°	-2.18	3.28	0.00	6.46	5.29	1.95	387.66	377.53	160.00	120.00
144°	-2.50	3.25	-0.01	5.23	6.93	1.88	405.22	411.72	160.00	120.00
162°	-2.82	3.30	0.04	3.18	7.63	1.86	391.67	390.52	160.00	120.00
180°	-3.12	3.27	0.04	1.35	8.61	1.79	405.18	410.89	160.00	120.00

Table B.2: Camera calibration of the CASIA database [55]. 'Yaw' is rotation respect to Y direction, 'Pitch' is rotation respect to X direction and 'Roll' is rotation respect to Z direction. T_x, T_y, T_z are translation distances, (F_x, F_y) is focal length and (C_x, C_y) is camera centre.

Appendix C

Frontal Recogntion GUI

C.1 Introduction

As a part of our research, Matlab-based GUI for the frontal-based gait recognition has been developed and is freely available upon request for further development and research. Currently the GUI facilitates depth image capturing from Kinect, Segmentation of depth Silhouette, online & offline training and recognition. The overall functionality of the developed GUI is illustrated in Figure C.1 and Figure C.2 shows the main interface of the GUI.

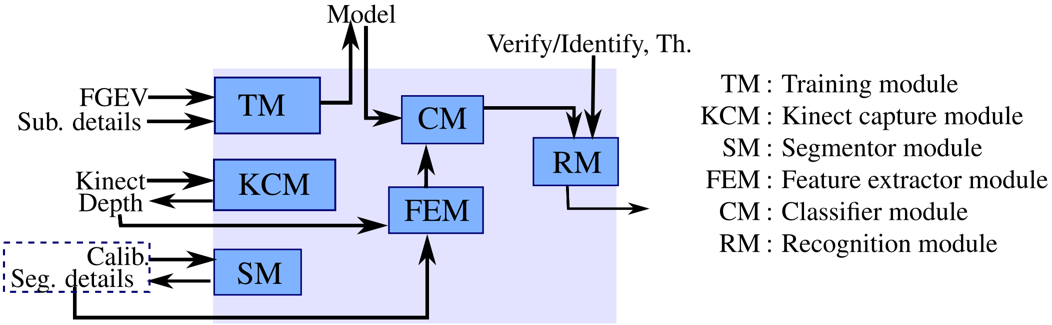
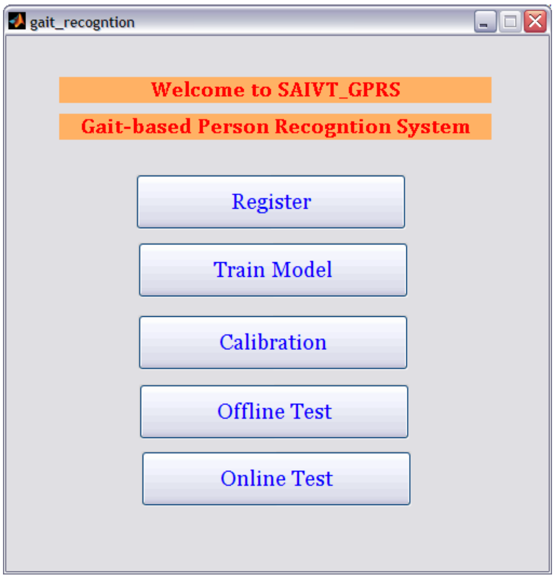


Figure C.1: Functionality of the frontal gait recognition GUI.

Functionality of each module in the GUI framework are explained in remaining of this section.

C.2 Training Module (TM)

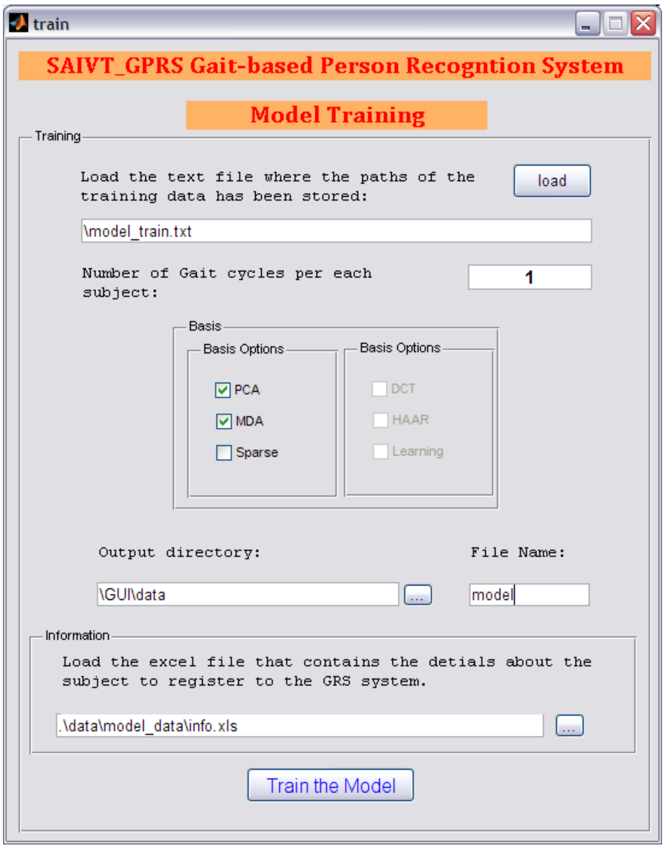
The training module works to form the underlying functionality for the GUI. I organised the preliminary data for the system to make the identity/ verification is for basis



Welcome to SAIVT_GPRS

Gait-based Person Recogntion System

Figure C.2: Main window of the frontal gait recognition GUI.



SAIVT_GPRS Gait-based Person Recogntion System

Model Training

Training

```
Load the text file where the paths of the
training data has been stored:
```

\model_train.txt

```
Number of Gait cycles per each          1
subject:
```

Basis

Basis Options	Basis Options
☑ PCA	□ DCT
☑ MDA	□ HAAR
□ Sparse	□ Learning

```
Output directory:                  File Name:
```

\GUI\data   model

Information

```
Load the excel file that contains the detials about the
subject to register to the GRS system.
```

.\data\model_data\info.xls

Figure C.3: Training module of the frontal gait recognition GUI.

learning. This takes Frontal GEV features of the training subjects and learns the basis from them. The required basis is optional input from the user from PCA, MDA and sparse with DCT, HAAR and Learned basis. Output from the TM is a structure that contains all the training subjects' transformed features on the learned basis, subject labels, selected basis option and personal details that need to be retrieved on verification/ identification. Figure C.3 shows the GUI interface of the training module.

C.3 Kinect Capture Module (KCM)

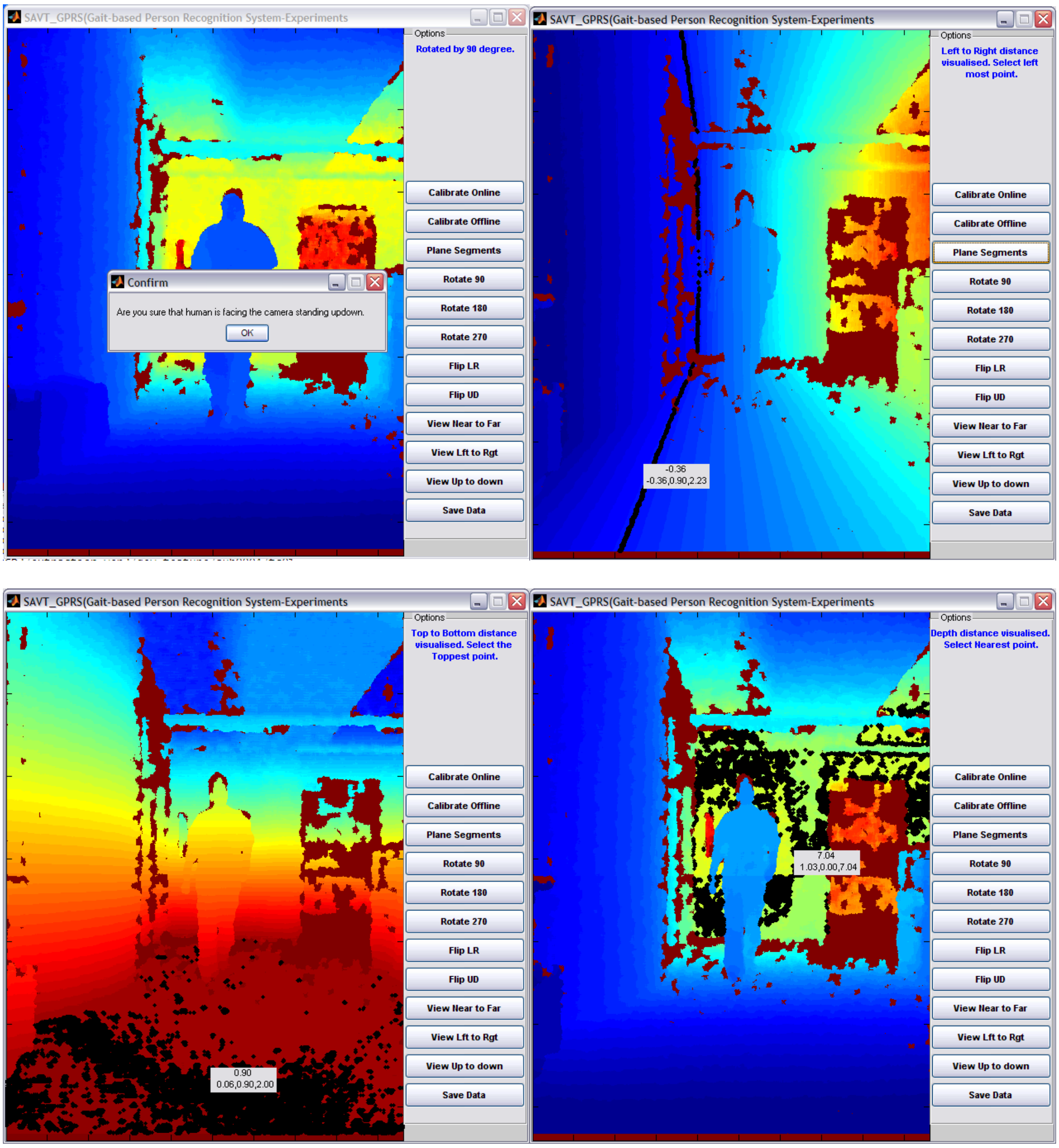
KCM enables the depth image capturing from the Kinect. It incorporates C++ wrapper to get the binary stream from Kinect and save the Depth, RGB image in 30fps. This module also enables change of quality of the storing image with different image format. However the capturing time is limited with the manually defined buffer size that is based on the computer's memory size.

C.4 Silhouette Segmentation Module (SSM)

A depth silhouette needs to be extracted from the depth image to create the frontal voxel volume. SSM visualises the depth to the user and gets the segmentation planes from the GUI interface and stores them. Figure C.4 illustrates the plane segmentation with appropriate visualisation of world coordinates in each axis direction.

C.5 Feature Extractor Module (FEM)

FEM does the main role as it extracts the FGEV feature from the straight depth raw images. It takes depth images from KCM for online testing or from stored data for offline testing. Calibration parameters and segmentation details stored/computed via SSM are used to extract the depth silhouette and transform that to 3D world coordinates. Then voxel volumes are constructed, gait cycles are identified and FGEVs are produced as explained in Section 5.2. Computed features are then transformed to the learned basis domain that is outputted from the training module.


SAVT_GPRS(Gait-based Person Recognition System-Experiments

Options

Rotated by 90 degree.

Confirm

Are you sure that human is facing the camera standing updown.

Left to Right distance visualised. Select left most point.

-0.36
-0.36,0.90,2.23

Top to Bottom distance visualised. Select the Toppest point.

0.90
0.06,0.90,2.00

Depth distance visualised. Select Nearest point.

7.04
1.03,0.00,7.04

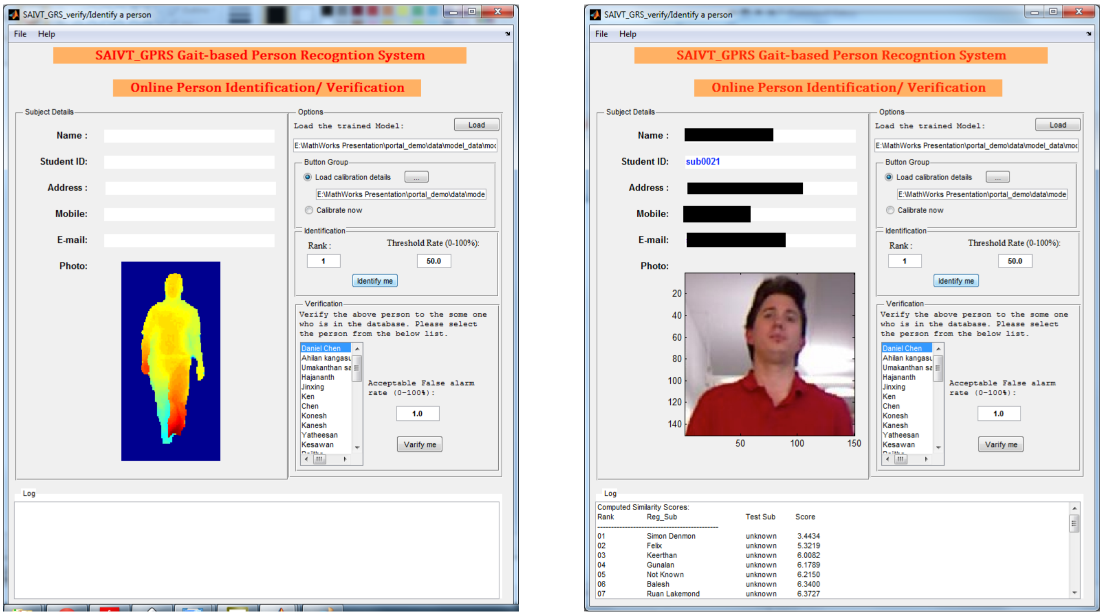
Figure C.4: Segmentation of planes using the frontal recognition GUI.

C.6 Classifier Module (CM)

Distance-to-feature vector of the testing subject and the enrolled subjects in the learned basis domain are computed using SRC-based classifier in classifier module.

C.7 Recognition Module (RM)

Based on the distances from CM, if the user requests to verify the testing subjects for the particular enrolled subject, the recognition module produces the verification results for the user input threshold false alarm rate. If the testing subject matches within the false alarm threshold, details for the particular subject are retrieved from the outputted model from TM. Similar to verification, if the user claims the identity for the particular rank, then identification of the subject with the enrolled subjects have been performed and the details of the identified subject with the given rank will be retrieved as shown in Figure C.5


SAIVT_GRS_verify/Identify a person
File
Help
SAIVT_GPRS Gait-based Person Recogntion System
Online Person Identification/ Verification
Subject Details
Photo:
Options
Load
Button Group
...
Identification
Identify me
Verification
Verify the above person to the some one who is in the database. Please select the person from the below list.
Varify me
Log
20
40
60
80
100
120
140
50
100
150

Name :
Student ID:
Address :
Mobile:
E-mail:

Load the trained Model: E:\MathWorks Presentation\portal_demo\data\model_data\moc

● Load calibration details E:\MathWorks Presentation\portal_demo\data\mode

○ Calibrate now

Rank : 1

Threshold Rate (0-100%): 50.0

Daniel Chen
Ahilan kangasu
Umakanthan sa
Hajananth
Jinxing
Ken
Chen
Konesh
Kanesh
Yatheesan
Kesawan

Acceptable False alarm rate (0-100%): 1.0

Name : [redacted]
Student ID: sub0021
Address : [redacted]
Mobile: [redacted]
E-mail: [redacted]

```
Computed Similarity Scores:
Rank        Reg_Sub                 Test Sub     Score
------------------------------------------
01          Simon Denmon            unknown      3.4434
02          Felix                   unknown      5.3219
03          Keerthan                unknown      6.0082
04          Gunalan                 unknown      6.1789
05          Not Known               unknown      6.2150
06          Balesh                  unknown      6.3400
07          Ruan Lakemond           unknown      6.3727
```

Figure C.5: Functionality of feature extraction and recognition modules. In first phase, gait energy volume-based features are extracted. Secondly, these features are matched with the enrolled subjects' features and similarity scores with meta data of the identified person are visualised.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 4–20, 2004.

[2] A. Lagorio, M. Tistarelli, M. Cadoni, C. B. Fookes, and S. Sridharan, "Liveness detection based on 3D face shape analysis," in *International Workshop on Biometrics and Forensics (IWBF).* Lisbon, Portugal: IEEE, 2013, pp. 1–4.

[3] A. Maeder, C. Fookes, and S. Sridharan, "Gaze based user authentication for personal computer applications," in *Intelligent Multimedia, Video and Speech Processing. Proceedings of 2004 International Symposium on.* IEEE, 2004, pp. 727–730.

[4] C. Fookes, A. Maeder, S. Sridharan, and G. Mamic, "Gaze based personal identification," *Behavioral Biometrics for Human Identification Intelligent Applications, IGI Global*, pp. 237–263, 2010.

[5] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2021–2040, 2003.

[6] "Fingerprint biometric," Digital image, http://www.freeimages.com/photo/227873, [Accessed: 2014-08-17].

[7] "Person recognition based on human iris," Digital image, http://www.iosrjournals.org/iosr-jce/full-issue/vol10-issue1.pdf, [Accessed: 2014-08-17].

[8] "Face recognition," Digital image, http://www.laymanpsychology.com/facial-recognition-psychology/, [Accessed: 2014-08-17].

[9] "Biometric security measures using hand," Digital image, http://blogs.swa-jkt.com/swa/10472/2012/08/16/biometric-security-measures/, [Accessed: 2014-08-17].

[10] “Forensic analysis using palm print,” Digital image, http://www.pinterest.com/jkacheroski/forensic-science/, [Accessed: 2014-08-17].

[11] “Identification using finger vein pattern,” Digital image, http://www.ind-safety.com/editorial/2013/04/IS1304_Tech-Update_01.html, [Accessed: 2014-08-17].

[12] “Gait recognition for identifying human at distance,” Digital image, http://gtresearchnews.gatech.edu/newsrelease/GAIT.htm, [Accessed: 2014-08-17].

[13] “Digital voice signal,” Digital image, http://www.123rf.com/photo_3488813_green-digital-wave.html, [Accessed: 2014-08-17].

[14] “Signature,” Digital image, http://en.wikipedia.org/wiki/Signature_recognition, [Accessed: 2014-08-17].

[15] “Keyboard stroke,” Digital image, http://thefinishedcopy.com/2014/03/30/how-i-wrote-1700-words-in-one-hour-using-this-simple-writing-system/, [Accessed: 2014-08-17].

[16] A. Ross and A. Jain, “Human recognition using biometrics: an overview,” *Annales Des Tlcommunications*, vol. 62, no. 1-2, pp. 11–35, 2007.

[17] Y. Sun, C. B. Fookes, N. Poh, and M. Tistarelli, “Cohort normalization based sparse representation for undersampled face recognition,” in *The 11th Asian Conference on Computer Vision*, K. Lee, Y. Matsushita, J. Rehg, and Z. Hu, Eds. Korea: Springer, Mar. 2013. [Online]. Available: http://eprints.qut.edu.au/58516/

[18] C. McCool, V. Chandran, S. Sridharan, and C. Fookes, “3D face verification using a free-parts approach,” *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1190–1196, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865508000329

[19] K. N. Thanh, C. B. Fookes, S. Sridharan, and S. Denman, “Focus-score weighted super-resolution for uncooperative iris recognition at a distance and on the move,” in *25th International Conference of Image and Vision Computing*, Queenstown, New Zealand, Nov. 2010. [Online]. Available: http://eprints.qut.edu.au/38118/

[20] C. Fookes, G. Mamic, C. McCool, and S. Sridharan, “Normalisation and recognition of 3D face data using robust hausdorff metric,” in *Digital Image Computing: Techniques and Applications (DICTA), 2008*, Dec. 2008, pp. 124–129.

[21] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, “Quality-driven super-resolution for less constrained iris recognition at a distance and on the move,” *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 4, pp. 1248–1258, Dec. 2011.

[22] C. B. Fookes, D. Chen, R. Lakemond, and S. Sridharan, “Robust facial feature extraction and matching,” *Journal of Pattern Recognition Research*, vol. 7, no. 1, pp. 140–154, 2012. [Online]. Available: http://eprints.qut.edu.au/52634/

[23] D. Chen, G. Mamic, C. B. Fookes, and S. Sridharan, “Scale-space volume descriptors for automatic 3D facial feature extraction,” *International Journal of Signal Processing*, vol. 5, no. 4, pp. 264–269, 2009. [Online]. Available: http://eprints.qut.edu.au/20497/

[24] F. Lin, C. Fookes, V. Chandran, and S. Sridharan, “Super-resolved faces for improved face recognition from surveillance video,” in *Advances in Biometrics*. Springer, 2007, pp. 1–10.

[25] K. Nguyen Thanh, C. B. Fookes, S. Sridharan, and S. Denman, “Feature-domain super-resolution for IRis recognition,” in *Proceedings of The 18th International Conference on Image Processing ICIP 2011*. IEEE, 2011.

[26] Fookes, Clinton and Lin, Frank and Chandran, Vinod and Sridharan, Sridha, “Evaluation of image resolution and super-resolution on face recognition performance,” *Journal of Visual Communication and Image Representation*, vol. 23, no. 1, pp. 75–93, 2012.

[27] K. Nguyen, S. Sridharan, S. Denman, and C. Fookes, “Feature-domain super-resolution framework for Gabor-based face and iris recognition,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2642–2649.

[28] K. Nguyen Thanh, C. B. Fookes, S. Sridharan, and S. Denman, “Feature-domain super-resolution for IRis recognition,” in *Proceedings of The 18th International Conference on Image Processing ICIP 2011*. IEEE, 2011.

[29] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*. IEEE, 2010, pp. 320–327.

[30] N. Boulgouris, D. Hatzinakos, and K. Plataniotis, "Gait recognition: a challenging signal processing technology for biometric identification," *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 78–90, 2005.

[31] R. Gross and J. Shi, "The Carnegie Mellon University (CMU) Motion of Body (MoBo) Database," Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-18, Jun. 2001.

[32] "Occluded legs," Digital image, http://www.englandathletics.org/england-athletics-news/england-athletics-international-race-walk-team-opportunities-2011, [Accessed: 2014-08-17].

[33] "The four bombers captured on cctv," Digital image, http://en.wikipedia.org/wiki/7_July_2005_London_bombings, [Accessed: 2014-08-17].

[34] "Walking on different surfaces," Digital image, http://u3alogan.blogspot.com.au/2013/01/walking-for-fitness-group.html, [Accessed: 2014-08-17].

[35] "Walking while carrying goods," Digital image, http://www.crimsonhexagon.com/business-intelligence/social-analytics/brand-affinity/five-social-personas-of-back-to-school-shoppers, [Accessed: 2014-08-17].

[36] "Walking with different clothing," Digital image, http://www.hawaiiweddinglimousine.com/oahu_honeymoon_tour_limousine.html, [Accessed: 2014-08-17].

[37] "Gait abnormalities," Digital image, http://neurologycoffecup.wordpress.com/2008/09/02/297/, [Accessed: 2014-08-17].

[38] "The ministry of silly walks," Digital image, http://blog.chocovenyl.co.uk/tag/monty-python-sketch/, [Accessed: 2014-08-17].

[39] "Changes in gait patterns with aging," Digital image, http://www.visualphotos.com/image/1x3743563/gait_patterns_illustration_depicting_the_gait, [Accessed: 2014-08-17].

[40] J. E. Boyd and J. J. Little, “Biometric gait recognition,” *Advanced Studies in Biometrics*, pp. 19–42, 2005.

[41] C. BenAbdelkader, R. Cutler, and L. Davis, “Person identification using automatic height and stride estimation,” in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 4. IEEE, 2002, pp. 377–380.

[42] I. Bouchrika and M. S. Nixon, “Model-based feature extraction for gait analysis and recognition,” in *Computer vision/computer graphics collaboration techniques*. Springer, 2007, pp. 150–160.

[43] M. S. Nixon, J. N. Carter, J. M. Nash, P. S. Huang, D. Cunado, and S. V. Stevenage, “Automatic gait recognition,” in *Motion Analysis and Tracking (Ref. No. 1999/103), IEE Colloquium on*, 1999, pp. 3/1–3/6.

[44] M. Nixon, “Gait biometrics,” *Biometric Technology Today*, vol. 16, no. 7-8, pp. 8–9, 2008.

[45] P. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, “Baseline results for the challenge problem of humanid using gait analysis,” in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 2002, pp. 130–135.

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

[47] M. S. Nixon, A. S. Aguado, and E. Montiel, “Invariant characterisation of the hough transform for pose estimation of arbitrary shapes,” *Pattern Recognition*, vol. 35, no. 5, pp. 1083–1097, 2002.

[48] M. S. Nixon, J. N. Carter, J. D. Shutler, and M. G. Grant, “New advances in automatic gait recognition,” *Information Security Technical Report*, vol. 7, no. 4, pp. 23–35, 2002.

[49] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi, “Semi-supervised intelligent surveillance system for secure environments,” in *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 2815–2820.

[50] S. Denman, M. Halstead, A. Bialkowski, C. B. Fookes, and S. Sridharan, “Can you describe him for me? a technique for semantic person search in video,” *Proceedings of Digital Image Computing: Techniques and Applications 2012*, pp. 1–8, 2012.

[51] D. L. Donoho, “Compressed sensing,” *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[52] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” in *IEEE Trans. on PAMI*, 2008.

[53] E. Candès and M. Wakin, “An introduction to compressive sampling,” *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[54] O. Concha, R. Xu, and M. Piccardi, “Compressive sensing of time series for human action recognition,” in *Proc. IEEE Int. Conf. on Digital Image Computing: Techniques and Applications (DICTA)*, Dec. 2010, pp. 454–461.

[55] “Chinese Academy of Sciences (CASIA) gait database, 2009,” http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp, [Accessed: 2014-08-17].

[56] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. B. Fookes, “The backfilled gei: a cross-capture modality gait feature for frontal and side-view gait recognition,” in *Proceedings of the 2012 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2012, pp. 1–8.

[57] M. S. Nixon, “Model-based gait recognition,” *Encyclopedia of Biometrics*, 2007.

[58] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 1297–1304.

[59] M. Hofmann and G. Rigoll, “Improved gait recognition using radient histogram energy image,” in *Proc. IEEE Int. Conf. on ICIP*, 2012, pp. 1389–1392.

[60] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[61] T. Jabid, M. Kabir, and O. Chae, “LDP for face recognition,” in *Proc. IEEE Int. Conf. on ICCE*, Jan. 2010, pp. 329–330.

[62] M. S. Nixon, D. Cunado, and J. N. Carter, “Automatic extraction and description of human gait models for recognition purposes,” *Computer Vision and Image Understanding*, vol. 90, no. 1, pp. 1–41, 2003.

[63] D. A. Winter, “Biomechanical motor patterns in normal walking,” *Journal of motor behavior*, vol. 15, no. 4, pp. 302–330, 1983.

[64] “Definition of Gait Cycle,” http://www.laboratorium.dist.unige.it/~piero/Teaching/Gait, [Accessed: 2014-08-17].

[65] “Defence Advanced Research Projects Agency (DARPA)- Human identification at a distance,” http://www.darpa.mil/, [Accessed: 2014-08-17].

[66] “University of Southampton (SOTON) - Database for Human Id at Distance, 2005,” http://www.gait.ecs.soton.ac.uk/database, [Accessed: 2014-08-17].

[67] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, “The humanid gait challenge problem: data sets, performance, and analysis,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 2, pp. 162–177, 2005.

[68] “University of Maryland Database (UMD), 2002,” http://www.umiacs.umd.edu/labs/pirl/hid/data.html, [Accessed: 2014-08-17].

[69] “Carnegie Mellon University (CMU) Graphics Lab Motion Database, 2003,” http://mocap.cs.cmu.edu/, accessed: 2014-08-17.

[70] “Georgia Tech. Database (GTD), 2001,” http://www.cc.gatech.edu/cpl/projects/hid/images.html, [Accessed: 2014-08-17].

[71] “Massachusetts Institute of Technology (MIT) Database, 2002,” http://www.ai.mit.edu/people/llee/HID/data.htm, [Accessed: 2014-08-17].

[72] M. S. Nixon and J. N. Carter, “Automatic recognition by gait,” *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.

[73] S. Cheung and C. Kamath, “Robust background subtraction with foreground validation for urban traffic video,” *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 2330–2340, 2005.

[74] S. Denman, C. Fookes, S. Sridharan, and D. Ryan, “Multi-modal object tracking using dynamic performance metrics,” in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*. IEEE, 2010, pp. 286–293.

[75] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, “Textures of optical flow for real-time anomaly detection in crowds,” in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*. IEEE, 2011, pp. 230–235.

[76] S. Denman, C. Fookes, and S. Sridharan, “Improved simultaneous computation of motion detection and optical flow for object tracking,” in *Digital Image Computing: Techniques and Applications, 2009. DICTA’09.* IEEE, 2009, pp. 175–182.

[77] Cucchiara, R. and Grana, C. and Piccardi, M. and Prati, A., “Detecting moving objects, ghosts, and shadows in video streams,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 10, pp. 1337–1342, 2003.

[78] B. P. L. Lo and S. A. Velastin, “Automatic congestion detection system for underground platforms,” in *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, 2001, pp. 158–161.

[79] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: real-time tracking of the human body,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 780–785, 1997.

[80] C. Stauffer and W. Grimson, “Adaptive background mixture models for real-time tracking,” *CVPR*, 1999.

[81] A. M. Elgammal, D. Harwood, and L. S. Davis, “Non-parametric model for background subtraction,” in *Proceedings of the 6th European Conference on Computer Vision-Part II.* Springer-Verlag, 2000, pp. 751–767.

[82] H. Bohyung, “Sequential kernel density approximation and its application to real-time visual tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1186–1197, 2008.

[83] D. Chen, S. Denman, and C. Fookes, “Accurate silhouette segmentation using motion detection and graph cuts,” in *Information Sciences Signal Processing*

and their Applications (ISSPA), 2010 10th International Conference on, 2010, pp. 81–84.

[84] D. Chen, S. Denman, C. Fookes, and S. Sridharan, "Accurate silhouettes for surveillance - improved motion segmentation using graph cuts," in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, 2010, pp. 369–374.

[85] D. Chen, B. Chen, G. Mamic, C. Fookes, and S. Sridharan, "Improved grabcut segmentation via GMm optimisation," in *Digital Image Computing: Techniques and Applications (DICTA), 2008*, Dec. 2008, pp. 39–45.

[86] D. Chen and C. Fookes, "Labelled silhouettes for human pose estimation," in *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*, May 2010, pp. 574–577.

[87] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The gait identification challenge problem: data sets and baseline algorithm," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 1, 2002, pp. 385–388vol.1.

[88] R. T. Collins, R. Gross, and S. Jianbo, "Silhouette-based human identification from body shape and gait," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 2002, pp. 366–371.

[89] J. P. Foster, M. S. Nixon, and A. Prügel-Bennett, "Automatic gait recognition using area-based metrics," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2489–2497, 2003.

[90] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic gait recognition by symmetry analysis," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2175–2183, 2003.

[91] J. D. Shutler and M. S. Nixon, "Zernike velocity moments for description and recognition of moving shapes," in *Proc. BMVC 2001*, 2001, pp. 705–714.

[92] W. Liang, T. Tieniu, H. Weiming, and N. Huazhong, "Automatic gait recognition based on statistical shape analysis," *Image Processing, IEEE Transactions on*, vol. 12, no. 9, pp. 1120–1131, 2003.

[93] Y. Zhang, N. Yang, W. Li, X. Wu, and Q. Ruan, “Gait recognition using procrustes shape analysis and shape context,” in *Computer Vision ACCV 2009*, ser. Lecture Notes in Computer Science, H. Zha, R.-i. Taniguchi, and S. Maybank, Eds. Springer Berlin / Heidelberg, 2010, vol. 5996, pp. 256–265.

[94] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa, “Identification of humans using gait,” *Image Processing, IEEE Transactions on*, vol. 13, no. 9, pp. 1163–1173, 2004.

[95] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, “Eigengait: Motion-based recognition of people using image self-similarity,” in *Audio- and Video-Based Biometric Person Authentication*, ser. Lecture Notes in Computer Science, J. Bigun and F. Smeraldi, Eds., vol. 2091. Springer Berlin / Heidelberg, 2001, pp. 284–294.

[96] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR ’91., IEEE Computer Society Conference on*, 1991, pp. 586–591.

[97] J. Little and J. Boyd, “Recognizing people by their gait: the shape of motion,” *Videre, Vol. 1, No. 2*, 1998.

[98] J. Han and B. Bhanu, “Individual recognition using gait energy image,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 2, pp. 316–322, 2006.

[99] X. tao Chen, Z. hui Fan, H. Wang, and Z. qing Li, “Automatic gait recognition using kernel principal component analysis,” in *Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on*, 2010, pp. 1–4.

[100] M. Qinyong, W. Shenkang, N. Dongdong, and Q. Jianfeng, “Gait recognition at a distance based on energy deviation image,” in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, 2007, pp. 621–624.

[101] C. Lin and K. Wang, “A behavior classification based on enhanced gait energy image,” in *Networking and Digital Society (ICNDS), 2010 2nd International Conference on*, vol. 2, 2010, pp. 589–592.

[102] J. King, “Gait analysis. an introduction,” *Journal of the Royal Society of Medicine*, vol. 85, no. 1, p. 62, 1992.

[103] D. Cunado, M. S. Nixon, and J. N. Carter, “Automatic extraction and description of human gait models for recognition purposes,” *Comput. Vis. Image Underst.*, vol. 90, no. 1, pp. 1–41, 2003.

[104] L. Lee and W. E. L. Grimson, “Gait analysis for recognition and classification,” in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 2002, pp. 148–155.

[105] M. S. Nixon, J. N. Carter, and C. Yam, “Automated person recognition by walking and running via model-based approaches,” *Pattern Recognition*, vol. 37, no. 5, pp. 1057–1072, 2004.

[106] L. Wang, H. Ning, T. Tan, and W. Hu, “Fusion of static and dynamic body biometrics for gait recognition,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, pp. 1449–1454vol.2.

[107] D. K. Wagg and M. S. Nixon, “Automated model-based extraction and analysis of gait,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 11–16.

[108] D. Wagg and M. Nixon, “Model-based gait enrolment in real-world imagery,” in *Proc. Multimodal User Authentication*. University of California, Santa Barbara, 2003, pp. 189–195.

[109] Z. Ziheng, A. Prugel-Bennett, and R. I. Damper, “A Bayesian framework for extracting human gait using strong prior knowledge,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 11, pp. 1738–1752, 2006.

[110] R. Urtasun and P. Fua, “3D tracking for gait characterization and recognition,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 17–22.

[111] Z. Guoying, L. Guoyi, L. Hua, and M. Pietikainen, “3D gait recognition using multiple cameras,” in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 529–534.

[112] M. Tong, Y. Liu, and T. S. Huang, "3D human model and joint parameter estimation from monocular image," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 797–805, 2007.

[113] A. Sherstyuk, "Kernel functions in convolution surfaces: a comparative analysis," *The Visual Computer*, vol. 15, no. 4, pp. 171–182, 1999.

[114] L. Haitao, C. Yang, and W. Zengfu, "A novel algorithm of gait recognition," in *Wireless Communications & Signal Processing, 2009. WCSP 2009. International Conference on*, 2009, pp. 1–5.

[115] I. Jolliffe, "Principal component analysis," *Springer Series in Statistics, Berlin: Springer, 1986*, vol. 1, 1986.

[116] R. Sundaram, "Multiple discriminant analysis," *Department of Electrical and Computer Engineering Mississippi State University Mississippi State, http//: www. isip. msstate. edu/publications/courses/ece_8443/papers/2001/mmda/p02_paper_v0. pdf*, 2001.

[117] X. Huang and N. V. Boulgouris, "Gait recognition using linear discriminant analysis with artificial walking conditions," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2010, pp. 2461–2464.

[118] S. Cotter, "Recognition of occluded facial expressions using a fusion of localized sparse representation classifiers," in *Proc. IEEE Int. Workshop on Digital Signal Processing and Signal Processing Education (DSP/SPE)*, Jan. 2011, pp. 437–442.

[119] G. Casella and R. L. Berger, *Statistical Inference*. Duxbury Press Belmont, CA, 1990, vol. 70.

[120] G. Veres, L. Gordon, J. Carter, and M. Nixon, "What image information is important in silhouette-based gait recognition?" in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, Jun. 2004, pp. II–776–II–782 Vol.2.

[121] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1505–1518, 2003.

[122] C. B. Fookes, J. A. Cook, S. Sridharan, and M. Tistarelli, "Frequency decomposition techniques for increased discriminative 3D facial information capture," in *Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission*, 2010.

[123] F. Wang, S. Wen, C. Wu, Y. Zhang, and H. Wang, "Gait recognition based on the fast Fourier transform and SVM," in *Control and Decision Conference (CCDC), 2011 Chinese*, May 2011, pp. 1091–1094.

[124] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "Gait recognition using linear time normalization," *Pattern Recognition*, vol. 39, no. 5, pp. 969–979, 2006.

[125] N. Suo, X. Qian, and J. Zhao, "Gait recognition based on kpca and knn," in *Environmental Science and Information Application Technology (ESIAT), 2010 International Conference on*, vol. 3, 2010, pp. 432–435.

[126] Y. Li, Y. Yin, L. Liu, S. Pang, and Q. Yu, "Semi-supervised gait recognition based on self-training," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, Sep. 2012, pp. 288–293.

[127] Y. ChewYean, M. S. Nixon, and J. N. Carter, "On the relationship of human walking and running: automatic person identification by gait," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 1, 2002, pp. 287–290vol.1.

[128] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 1055–1064, 1999.

[129] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 148–155.

[130] R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 5, pp. 828–838, 2005.

[131] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, Dec. 1966.

[132] L. E. Baum and J. A. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bulletin of the American Mathematical Society*, vol. 73, no. 3, pp. 360–363, May 1967. [Online]. Available: http://projecteuclid.org/euclid.bams/1183528841

[133] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171, Feb. 1970.

[134] P. M. Baggenstoss, "A modified baum-welch algorithm for hidden markov models with multiple observation spaces," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 4, pp. 411–416, 2001.

[135] J. Forney, G. D., "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[136] S. Austin, R. Schwartz, and P. Placeway, "The forward-backward search algorithm," in *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*, 1991, pp. 697–700vol. 1.

[137] R. Chellappa, A. K. Roy-Chowdhury, and A. Kale, "Human identification using gait and face omitted," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007, pp. 1–2.

[138] M. S. Nixon, M. Goffredo, and J. N. Carter, "Front-view gait recognition," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, 2008, pp. 1–6.

[139] M. Butt, O. Henniger, A. Nouak, and A. Kuijper, "Privacy protection of biometric templates," in *HCI International 2014-Posters Extended Abstracts*. Springer, 2014, pp. 153–158.

[140] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image," in

Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, Sep. 2009, pp. 1058–1064.

[141] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. Int. Conf. on ICPR*, vol. 4, 0-0 2006, pp. 441–444.

[142] J. W. Davis, "Hierarchical motion history images for recognizing human motion," in *Proc. IEEE Workshop on Detection and Recognition of Events in Video*, 2001, pp. 39–46.

[143] Ho, Jeffrey and Yang, Ming-Hsuan and Lim, Jongwoo and Lee, Kuang-Chih and Kriegman, David, "Clustering appearances of objects under varying illumination conditions," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–11.

[144] D. Chen, P. Chou, C. B. Fookes, and S. Sridharan, "Multi-view human pose estimation using modified five-point skeleton model," in *International Conference on Signal Processing and Communication Systems 2007*, Gold Coast, Australia, 2008.

[145] P. K. Larsen, E. B. Simonsen, and N. Lynnerup, "Gait analysis in forensic medicine," *Journal of forensic sciences*, vol. 53, no. 5, pp. 1149–1153, 2008.

[146] "Burglar jailed over unusual walk," http://news.bbc.co.uk/2/hi/uk_news/england/lancashire/7343702.stm, [Accessed: 2014-08-17].

[147] T. Lee, M. Belkhatir, and S. Sanei, "A comprehensive review of past and present vision-based techniques for gait recognition," *Multimedia Tools and Applications*, pp. 1–37, 2013.

[148] "SmartGate for passport control," http://www.customs.gov.au/smartgate/default.asp, [Accessed: 2014-08-17].

[149] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[150] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang, "Chrono-gait image: A novel temporal template for gait recognition," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 257–270.

[151] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1598–1603, 2011.

[152] C. McCool, G. Mamic, C. Fookes, and S. Sridharan, "Normalisation of 3D face data," in *Proc. IEEE Int. Conf. on SPCS*, 2007, pp. 17–19.

[153] R. Basri, T. Hassner, and L. Zelnik-Manor, "Approximate nearest subspace search with applications to pattern recognition," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[154] S. Li and J. Lu, "Face recognition using the nearest feature line method," *Neural Networks, IEEE Transactions on*, vol. 10, no. 2, pp. 439–443, 1999.

[155] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, May 2008.

[156] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[157] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 2, pp. 218–233, 2003.

[158] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[159] D. Donoho and Y. Tsaig, "Fast solution of $\ell 1_1$ -norm minimization problems when the solution May be sparse," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 4789–4812, 2008.

[160] R. G. Baraniuk, "Compressive sensing [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 118–121, 2007.

[161] T. Jabid, M. Kabir, and O. Chae, “Gender classification using LDP,” in *Proc. IEEE Int. Conf. on ICPR*. IEEE Computer Society, 2010, pp. 2162–2165.

[162] J. Ryu and S. Kamata, “Front view gait recognition using spherical space model with human point clouds,” in *Proc. IEEE Int. Conf. on ICIP*, Sep. 2011, pp. 3209–3212.

[163] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O’toole, D. S. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, and S. Weimer, “Overview of the multiple biometrics grand challenge,” in *In Proc. Int. Conf. on Biometrics*, 2009, pp. 705–714.

[164] S. Yu, D. Tan, and T. Tan, “Modelling the effect of view angle variation on appearance-based gait recognition,” in *Computer Vision–ACCV 2006*. Springer, 2006, pp. 807–816.

[165] X. Huang and N. Boulgouris, “Gait recognition with shifted energy image and structural feature extraction,” *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 2256–2268, 2012.

[166] N. Liu and Y. Tan, “View invariant gait recognition,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1410–1413.

[167] R. Nevatia, “Camera calibration from video of a walking human,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 9, p. 1513, 2006.

[168] D. Zhang, *Automated biometrics: technnologies and systems*. Springer, 2000.

[169] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*. Springer, 1999.

[170] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, “3D ellipsoid fitting for multi-view gait recognition,” in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*. IEEE, 2011, pp. 355–360.

[171] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. B. Fookes, “Gait energy volumes and frontal gait recognition using depth images,” in *International*

Joint Conference on Biometrics. Washington DC, USA: IEEE, October 2011. [Online]. Available: http://eprints.qut.edu.au/46382/

[172] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "Histogram of weighted local directions for gait recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, 2013, pp. 125–130.

[173] Sivapalan, Sabesan and Rana, Rajib Kumar and Chen, Daniel and Sridharan, Sridha and Denmon, Simon and Fookes, Clinton, "Compressive sensing for gait recognition," in *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*. IEEE, 2011, pp. 567–571.

[174] Lu, Can-Yi and Huang, De-Shuang, "Optimized projections for sparse representation based classification," *Neurocomputing*, vol. 113, pp. 213–219, 2013.